

(12) PATENT
(19) AUSTRALIAN PATENT OFFICE

(11) Application No. AU 199748747 B2
(10) Patent No. 730900

(54) Title
Improvements in or relating to starch content of plants

(51)⁶ International Patent Classification(s)
C12N 015/82 C12N 009/10
A01H 005/00 C12N 015/54
C12N 001/21 C12Q 001/68

(21) Application No: 199748747

(22) Application Date: 1997 .11 .04

(87) WIPO No: W098/20145

(30) Priority Data

(31) Number (32) Date
9623095 1996 .11 .05

(33) Country
GB

(43) Publication Date: 1998 .05 .29

(43) Publication Journal Date: 1998 .07 .23

(44) Accepted Journal Date: 2001 .03 .15

(71) Applicant(s)
National Starch and Chemical Investment Holding Corporation

(72) Inventor(s)
Stephen Alan Jobling; Richard Safford

(74) Agent/Attorney
COLLISON and CO.GPO Box 2556,ADELAIDE SA 5001

(56) Related Art
PLANT MOLECULAR BIOLOGY. 20(5):809-819
THE PLANT JOURNAL. 7(1):3-5
PLANT MOLECULAR BIOLOGY 30:97-108

OPI DATE 29/05/98 APPLN. ID 48747/97
AGJP DATE 23/07/98 PCT NUMBER PCT/GB97/03032



AU9748747

(T)

(51) International Patent Classification 6: C12N 15/82, 9/10, C12Q 1/68, C12N 1/21, 15/54, A01H 5/00		A2	(11) International Publication Number: WO 98/20145
(21) International Application Number: PCT/GB97/03032		(43) International Publication Date: 14 May 1998 (14.05.98)	
(22) International Filing Date: 4 November 1997 (04.11.97)		(53) Designated States: AU, BR, CA, JP, KR, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).	
(30) Priority Data: 9623095.8 5 November 1996 (05.11.96) GB		Published Without international search report and to be republished upon receipt of that report.	
(71) Applicant (for all designated States except US): NATIONAL STARCH AND CHEMICAL INVESTMENT HOLDING CORPORATION (US/US); Suite 22, 601 Silverside Road Wilmington, DE 19809 (US). 3411 19809			
(72) Inventors; and (75) Inventors/Applicants (for US only): JOBLING, Stephen, Alan (GB/GB); 19 Burwell Road, Eaton Socon, Huntingdon PE19 3QQ (GB). SAFFORD, Richard (GB/GB); 10 Furness Close, Bedford, Bedfordshire MK41 8RN (GB).			
(74) Agent: KEITH W NASH & CO.; 90-92 Regent Street, Cambridge CB2 1DP (GB).			

(54) Title: IMPROVEMENTS IN OR RELATING TO STARCH CONTENT OF PLANTS

(57) Abstract

Disclosed is a nucleic acid sequence encoding a polypeptide having starch branching enzyme (SBE) activity, the encoded polypeptide comprising an effective portion of the amino acid sequence shown in Figure 4 or Figure 13.

Title: Improvements in or Relating to Starch Content of Plants**Field of the Invention**

This invention relates to novel nucleic acid sequences, vectors and host cells comprising the nucleic acid sequence(s), to polypeptides encoded thereby, and to a method of altering a host cell by introducing the nucleic acid sequence(s) of the invention.

Background to the Invention

Starch consists of two main polysaccharides, amylose and amylopectin. Amylose is a linear polymer containing α -1,4 linked glucose units, while amylopectin is a highly branched polymer consisting of a α -1,4 linked glucan backbone with α -1,6 linked glucan branches. In most plant storage reserves amylopectin constitutes about 75% of the starch content. Amylopectin is synthesized by the concerted action of soluble starch synthase and starch branching enzyme [α -1,4 glucan: α -1,4 glucan 6-glycosyltransferase, EC 2.4.1.18]. Starch branching enzyme (SBE) hydrolyses α -1,4 linkages and rejoins the cleaved glucan, via an α -1,6 linkage, to an acceptor chain to produce a branched structure. The physical properties of starch are strongly affected by the relative abundance of amylose and amylopectin, and SBE is therefore a crucial enzyme in determining both the quantity and quality of starches produced in plant systems.

Starches are commercially available from several plant sources including maize, potato and cassava. Each of these starches has unique physical characteristics and properties and a variety of possible industrial uses. In maize there are a number of naturally occurring mutants which have altered starch composition such as high amylopectin types ("waxy" starches) or high amylose starches but in potato and cassava no such mutants exist on a commercial basis as yet.

Genetic modification offers the possibility of obtaining new starches which may have novel and potentially useful characteristics. Most of the work to date has involved potato plants because they are amenable to genetic manipulation i.e. they can be transformed using *Agrobacterium* and regenerated easily from tissue culture. In addition many of the genes involved in starch biosynthesis have been cloned from potato and thus are available as targets for genetic manipulation, for example, by antisense inhibition of expression or sense suppression.

Cassava (*Manihot esculenta* L. Crantz) is an important crop in the tropics, where its starch-filled roots are used both as a food source and increasingly as a source of starch. Cassava is a high yielding perennial crop that can grow on poor soils and is also tolerant of drought. Cassava starch being a root-derived starch has properties similar but not identical to potato starch and is composed of 20-25% amylose and 75-80% amylopectin (Rickard *et al.*, 1991. *Trop. Sci.* 31, 189-207). Some of the genes involved in starch biosynthesis have been cloned from cassava, including starch branching enzyme I (SBE I) (Salehuzzaman *et al.*, 1994 *Plant Science* 98, 53-62), and granule bound starch synthase I (GBSS I) (Salehuzzaman *et al.*, 1993 *Plant Molecular Biology* 23, 947-962) and some work has been done on their expression patterns although only in *in vitro* grown plants (Salehuzzaman *et al.*, 1994 *Plant Science* 98, 53-62).

In most plants studied to date e.g. maize (Boyer & Preiss, 1978 *Biochem. Biophys. Res. Comm.* 80, 169-175), rice (Smyth, 1988 *Plant Sci.* 57, 1-8) and pea (Smith, *Planta* 175, 270-279), two forms of SBE have been identified, each encoded by a separate gene. A recent review by Burton *et al.* (1995 *The Plant Journal* 7, 3-15) has demonstrated that the two forms of SBE constitute distinct classes of the enzyme such that, in general, enzymes of the same class from different plants may exhibit greater similarity than enzymes of different classes from the same plant. In their review, Burton *et al.* termed the two respective enzyme families class "A" and class "B", and the reader is referred thereto (and to the references cited therein) for a detailed discussion of the distinctions between the two classes. One general distinction of note would appear to be the presence, in class A SBE molecules, of a flexible N-terminal domain, which is not found in class B molecules. The distinctions noted by Burton *et al.* are relied on herein to define class A and class B SBE

molecules, which terms are to be interpreted accordingly.

Many organisations have interests in obtaining modified Cassava starches by means of genetic modification. This is impossible to achieve however, unless the plant is amenable to transformation and regeneration, and the starch biosynthesis genes which are to be targeted for modification must be cloned. The production of transgenic cassava plants has only recently been demonstrated (Taylor *et al.*, 1996 Nature Biotechnology 14, 726-730; Schöpke *et al.*, 1996 Nature Biotechnology 14, 731-735; and Li *et al.*, 1996 Nature Biotechnology 14, 736-740). The present invention concerns the identification, cloning and sequencing of a starch biosynthetic gene from Cassava, suitable as a target for genetic manipulation.

Summary of the Invention

In a first aspect the invention provides a nucleic acid sequence encoding a polypeptide having starch branching enzyme (SBE) activity, the polypeptide comprising an effective portion of the amino acid sequences shown in Figure 4 or Figure 13. The nucleic acid is conveniently in substantial isolation, especially in isolation from other naturally associated nucleic acid sequences.

An "effective portion" of the amino acid sequences may be defined as a portion which retains sufficient SBE activity when expressed in *E. coli* KV832 to complement the branching enzyme mutation therein. The amino acid sequences shown in Figures 4 and 13 include the N terminal transit peptide, which comprises about the first 50 amino acid residues. As those skilled in the art will be well aware, such a transit peptide is not essential for SBE activity. Thus the mature polypeptide, lacking a transit peptide, may be considered as one example of an effective portion of the amino acid sequence shown in Figure 4 or Figure 13.

Other effective portions may be obtained by effecting minor deletions in the amino acid sequence, whilst substantially preserving SBE activity. Comparison with known class A SBE sequences, with the benefit of the disclosure herein, will enable those skilled in the

art to identify regions of the polypeptide which are less well conserved and so amenable to minor deletion, or amino acid substitution (particularly, conservative amino acid substitution) whilst substantially preserving SBE activity. Such less well-conserved regions are generally found in the N terminal amino acid residues (up to the triple proline "elbow" at residues 138-140 in Figure 4 and up to the proline elbow at residues 143-145 in Figure 13) and in the last 50 residues or so of the C terminal, and in particular in the acidic tail of the C terminal.

Conveniently the nucleic acid sequence is obtainable from cassava, preferably obtained therefrom, and typically encodes a polypeptide obtainable from cassava. In a particular embodiment, the encoded polypeptide may have the amino acid sequence NSKH at about position 697 (in relation to Figure 4), which sequence appears peculiar to an isoform of the SBE class A enzyme of cassava, other class A SBE enzymes having the conserved sequence DA D/E Y (Burton *et al.*, 1995 cited above).

In a particular aspect of the invention there is provided a nucleic acid comprising a portion of nucleotides 21 to 2531 of the nucleic acid sequence shown in Figure 4, or a functionally equivalent nucleic acid sequence. Such functionally equivalent nucleic acid sequences include, but are not limited to, those sequences which encode substantially the same amino acid sequence but which differ in nucleotide sequence from that shown in Figure 4 by virtue of the degeneracy of the genetic code. For example, a nucleic acid sequence may be altered (e.g. "codon optimised") for expression in a host other than cassava, such that the nucleotide sequence differs substantially whilst the amino acid sequence of the encoded polypeptide is unchanged. Other functionally equivalent nucleic acid sequences are those which will hybridise under stringent hybridisation conditions (e.g. as described by Sambrook *et al.*, Molecular Cloning. A Laboratory Manual, CSH, i.e. washing with 0.1xSSC, 0.5% SDS at 68°C) with the sequence shown in Figure 4. Figure 10 shows a functionally equivalent sequence designated "125 + 94", which includes a region corresponding to the 3' coding portion of the sequence in Figure 4. Figure 13 shows a functionally equivalent sequence which comprises a second complete SBE coding sequence (the SBE-derived sequence is from nucleotides 35 to 2760, of which the coding sequence is nucleotides 131-2677, the rest of the sequence in the figure is vector-derived).

Functionally equivalent DNA sequences will preferably comprise at least 200-300bp, more preferably 300-600bp, and will exhibit at least 88% identity (more preferably at least 90%, and most preferably at least 95% identity) with the corresponding region of the DNA sequence shown in figures 4 or 10. Those skilled in the art will readily be able to conduct a sequence alignment between the putative functionally equivalent sequence and those detailed in Figures 4 or 10 - the identity of the two sequences is to be compared in those regions which are aligned by standard computer software, which aligns corresponding regions of the sequences.

In particular embodiments the nucleic acid sequence may alternatively comprise a 5' and/or a 3' untranslated region ("UTR"), examples of which are shown in Figures 2 and 4. Figure 9 includes a 3' UTR, as nucleotides 688-1044 and Figure 10 includes 3' UTR as nucleotides 1507-1900 (which nucleotides correspond to the first base after the "stop" codon to the base immediately preceding the poly (A) tail). Any one of the sequences defined above, or a functional equivalent thereof (as defined by hybridisation properties, as set out in the preceding paragraph), could be useful in sense or anti-sense inhibition of corresponding genes, as will be apparent to those skilled in the art. It will also be apparent to those skilled in the art that such regions may be modified so as to optimise expression in a particular type of host cell and that the 5' and/or 3' UTRs could be used in isolation, or in combination with a coding portion of the sequence of the invention. Similarly, a coding portion could be used without a 5' or a 3' UTR if desired.

In a further aspect, the invention provides a replicable nucleic acid construct comprising any one of the nucleic acid sequences defined above. The construct will typically comprise a selectable marker and may allow for expression of the nucleic acid sequence of the invention. Conveniently the vector will comprise a promoter (especially a promoter sequence operable in a plant and/or a promoter operable in a bacterial cell) and one or more regulatory signals known to those skilled in the art.

In another aspect the invention provides a polypeptide having SBE activity, the polypeptide comprising an effective portion of the amino acid sequence shown in Figure 4 or Figure 13. The polypeptide is conveniently one obtainable from cassava, although it may be

derived using recombinant DNA techniques. The polypeptide is preferably in substantial isolation from other polypeptides of plant origin, and more preferably in substantial isolation from any other polypeptides. The polypeptide may have amino acid residues NSKH at about position 697 (in the sequence shown in Figure 4), instead of the sequence DA D/E Y found in other SBE class A polypeptides. The polypeptide may be used in a method of modifying starch *in vitro*, the method comprising treating starch under suitable conditions (of temperature, pH etc.) with an effective amount of the polypeptide.

Those skilled in the art will appreciate that the disclosure of the present specification can be utilised in a number of ways. In particular, the characteristics of a host cell may be altered by recombinant DNA techniques. Thus, in a further aspect, there is provided a method by which a host cell may be altered by introduction of a nucleic acid sequence comprising at least 200bp and exhibiting at least 88% sequence identity (more preferably at least 90%, and most preferably at least 95% identity) with the corresponding region of the DNA sequence shown in Figures 4, 9, 10 or 13, operably linked in the sense or (preferably) in the anti-sense orientation to a suitable promoter active in the host cell, and causing transcription of the introduced nucleic acid sequence, said transcript and/or the translation product thereof being sufficient to interfere with the expression of a homologous gene naturally present in said host cell, which homologous gene encodes a polypeptide having SBE activity. The altered host cell is typically a plant cell, such as a cell of a cassava, banana, potato, sweet potato, tomato, pea, wheat, barley, oat, maize, or rice plant.

Desirably the method further comprises the introduction of one or more nucleic acid sequences which are effective in interfering with the expression of other homologous gene or genes naturally present in the host cell. Such other genes whose expression is inhibited may be involved in starch biosynthesis (e.g. an SBE I gene), or may be unrelated to SBE II.

Those skilled in the art will be aware that both anti-sense inhibition, and "sense suppression" of expression of genes, especially plant genes, has been demonstrated (e.g. Matzke & Matzke 1995 Plant Physiol. 107, 679-685).

It is believed that antisense methods are mainly operable by the production of antisense mRNA which hybridises to the sense mRNA, preventing its translation into functional polypeptide, possibly by causing the hybrid RNA to be degraded (e.g. Sheehy *et al.*, 1988 PNAS 85, 8805-8809; Van der Krol *et al.*, Mol. Gen. Genet. 220, 204-212). Sense suppression also requires homology between the introduced sequence and the target gene, but the exact mechanism is unclear. It is apparent however that, in relation to both antisense and sense suppression, neither a full length nucleotide sequence, nor a "native" sequence is essential. Preferably the nucleic acid sequence used in the method will comprise at least 200-300bp, more preferably at least 300-600bp, of the full length sequence, but by simple trial and error other fragments (smaller or larger) may be found which are functional in altering the characteristics of the plant. It is also known that untranslated portions of sequence can suffice to inhibit expression of the homologous gene - coding portions may be present within the introduced sequence, but they do not appear to be essential under all circumstances.

The inventors have discovered that there are at least two class A SBE genes in cassava. A fragment of a second gene has been isolated, which fragment directs the expression of the C terminal 481 amino acids of cassava class A SBE (see Figure 10) and comprises a 3' untranslated region. Subsequently, a complete clone of the second gene was also recovered (see Figure 12). The coding portions of the two genes show some slight differences, and the second SBE gene may be considered as functionally equivalent to the corresponding portion of the nucleotide sequence shown in Figure 4. However, the 3' untranslated regions of the two genes show marked differences. Thus the method of altering a host cell may comprise the use of a sufficient portion of either gene so as to inhibit the expression of the naturally occurring homologous gene. Conveniently, a portion of nucleotide sequence is employed which is conserved between both genes. Alternatively, sufficient portions of both genes may be employed, typically using a single construct to direct the transcription of both introduced sequences.

In addition, as explained above, it may be desired to cause inhibition of expression of the class B SBE (i.e. SBE I) in the same host cell. A number of class B SBE gene sequences are known, including portions of the cassava class B SBE (Salehuzzaman *et al.*, 1994

Plant Science 98, 53-62) and any one of these may prove suitable. Preferably the sequence used is that which derives from the host cell sought to be altered (e.g. when altering the characteristics of a cassava plant cell, it is generally preferred to use sense or anti-sense sequences corresponding exactly to at least portions of the cassava gene whose expression is sought to be inhibited).

In a further aspect the invention provides an altered host cell, into which has been introduced a nucleic acid sequence comprising at least 200bp and exhibiting at least 88% sequence identity (more preferably at least 90%, and most preferably at least 95% identity) with the corresponding region of the DNA sequence shown in Figures 4, 9, 10 or 13, operably linked in the sense or anti-sense orientation to a suitable promoter, said host cell comprising a natural gene sharing sequence homology with the introduced sequence.

The host cell may be a micro-organism (such as a bacterial, fungal or yeast cell) or a plant cell. Conveniently the host cell altered by the method is a cell of a cassava plant, or another plant with starch storage reserves, such as banana, potato, sweet potato, tomato, pea, wheat, barley, oat, maize, or rice plant. Typically the sequence will be introduced in a nucleic acid construct, by way of transformation, transduction, micro-injection or other method known to those skilled in the art. The invention also provides for a plant into which has been introduced a nucleic acid sequence of the invention, or the progeny of such a plant.

The altered plant cell will preferably be grown into an altered plant, using techniques of plant growth and cultivation well-known to those skilled in the art of re-generating plantlets from plant cells.

The invention also provides a method of obtaining starch from an altered plant, the plant being obtained by the method defined above. Starch may be extracted from the plant by any of the known techniques (e.g. milling). The invention further provides starch obtainable from a plant altered by the method defined above, the starch having altered properties compared to starch extracted from an equivalent but unaltered plant. Conveniently the altered starch is obtained from an altered plant selected from the group

consisting of cassava, potato, pea, tomato, maize, wheat, barley, oat, sweet potato and rice. Typically the altered starch will have increased amylose content.

The invention will now be further described by way of illustrative examples and with reference to the accompanying drawings, in which:-

Figure 1 is a schematic illustration of the cloning strategy for cassava SBE II. The top line represents the size of a full length clone with distances in kilobases (kb) and arrows representing oligonucleotides (rightward pointing arrows are sense strand, leftward are on opposite strand). The long thick arrow is the open reading frame with start and stop codons shown. Below this are shown the 3' RACE, 5' RACE and PCR clones identified either by the plasmid name (shown in brackets above the line) or the clone number (shown to the left of the clone) for the 5' RACE only. Also shown (by an x) in the 5' RACE clones are positions of small deletions or introns.

Figure 2 shows the DNA sequence and predicted ORF of csbe2con.seq. This sequence is a consensus of 3' RACE pSJ94 and 5' RACE clones 27/9, 11 and 28. The first 64 base pairs are derived from the RoRIdT17 adaptor primer/dT tail followed by the SBE sequence. The one long open reading frame is shown in one letter code below the double strand DNA sequence. Also shown is the upstream ORF (MQL...LPW).

Figure 3 shows an alignment of the 5' region of cassava SBE II csbe2con and pSJ99 (clones 20 and 35) DNA sequences. Differences from the consensus sequence are shaded.

Figure 4 shows the DNA sequence and predicted ORF of (full length) cassava SBE II tuber cDNA in pSJ107. The sequence shown is from the CSBE214 to the CSBE218 oligonucleotide. The DNA sequence is sequence ID No. 28 in the attached sequence listing; the amino acid sequence is Seq ID No. 29.

Figure 5 shows an alignment of 3' region of cassava SBE II pSJ116 and 125+94 DNA sequences. The top line is the 125 + 94 sequence and the bottom SJ116 sequence. Identical nucleotides are indicated by the same letter in the middle line. Differences are

indicated by a gap, and dashed lines indicate gaps introduced to optimise alignment.

Figure 6 shows an alignment of carboxy terminal region of pSJ116 and 125+94 protein sequences. The top sequence is from 125+94 and the bottom from pSJ116. Identical amino acid residues are shown with the same letter, conserved changes with a colon and neutral changes with a period.

Figure 7 shows a phylogenetic tree of starch branching enzyme proteins. The length of each pair of branches represents the distance between sequence pairs. The scale beneath the tree measures the distance between sequences (units indicate the number of substitution events). Dotted lines indicate a negative branch length because of averaging the tree. Zmcon12.pro is maize SBE II, psstb1.pro is pea SBE I (Bhattacharyya *et al* 1990 Cell 60, 115-121) and atsbe2-1 & 2-2.pro are two SBE II proteins from *Arabidopsis thaliana* (Fisher *et al* 1996 Plant Mol. Biol. 30, 97-108). SJ107.pro is representative of a cassava SBE II sequence, and potsbe2.pro is a potato SBE II sequence known to the inventors.

Figure 8 is an alignment of SBE II proteins. Protein sequences are indicated in one letter code. The top line represents the consensus sequence, below which is shown the consensus ruler and the individual SBE II sequences. Residues matching the consensus are shaded. Dashes represent gaps introduced to optimise alignment. Sequence identities are shown at the right of the figure and are as Figure 7, except that SJ107.pro is cassava SBE II.

Figure 9 shows the DNA sequence and predicted ORF of a cassava SBE II cDNA isolated by 3' RACE (plasmid pSJ 101).

Figure 10 shows the consensus DNA sequence and predicted ORF of a second cassava SBE II cDNA isolated by 3' and 5' RACE (sequence designated 125+94 is from plasmid pSJ125 and pSJ94, spliced at the CSBE217 oligo sequence).

Figure 11 is a schematic diagram of the plant transformation vector pSJ64. The black line represents the DNA sequence. The hashed line represents the bacterial plasmid backbone

(containing the origin of replication and bacterial selection marker) and is not shown in full. The filled triangles represent the T-DNA borders (RB = right border, LB = left border). Relevant restriction enzyme sites are shown above the black line with the approximate distances (in kilobases) between sites marked by an asterisk shown underneath. The thinnest arrows represent polyadenylation signals (pAnos = nopaline synthase, pAg7 = Agrobacterium gene 7), the intermediate arrows represent protein coding regions (SBE II = cassava SBE II, HYG = hygromycin resistance gene) and the thick arrows represent promoter regions (P-2x35S = double CaMV 35S promoter, P-nos = nopaline synthase promoter).

Figure 12 is a schematic illustration of the cloning strategy used to isolate a second cassava SBE II gene. The top line represents the size of a full length clone with distances in kilobases (kb) and arrows representing oligonucleotides (rightward pointing arrows are sense strand, leftward are on opposite strand). The long thick arrow is the open reading frame with start and stop codons shown. Below this are shown the 3' RACE, 5' RACE and PCR clones identified either by the plasmid name (shown in brackets above the line) or the clone number (shown to the right of the clone).

Figure 13 shows the DNA sequence and predicted ORF of a second full length cassava SBE II tuber cDNA in pSJ146. Nucleotides 35-2760 are SBE II sequence and the remainder are from the pT7Blue vector. The DNA sequence of Figure 13 is Seq ID No. 30, and the amino acid sequence is Seq ID No. 31, in the attached sequence listing.

Example 1

This example relates to the isolation and cloning of SBE II sequences from cassava.

Recombinant DNA manipulations

Standard procedures were performed essentially according to Sambrook *et al.* (1989 Molecular cloning A laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.). DNA sequencing was performed on an ABI automated DNA sequencer and sequences manipulated using DNASTAR software for the Macintosh.

Rapid Amplification of cDNA ends (RACE) and PCR conditions

5' and 3' RACE were performed essentially according to Frohman *et al.*, (1988 Proc. Natl. Acad. Sci. USA 85, 8998-9002) but with the following modifications.

For 3' RACE, 5 µg of total RNA was reverse transcribed using 5 pmol of the RACE adaptor RoRidT17 as primer and Stratascript RNase H- reverse transcriptase (50 U) in a 50 µl reaction according to the manufacturer's instructions (Stratagene). The reaction was incubated for 1 hour at 37°C and then diluted to 200 µl with TE (10 mM Tris HCl, 1 mM EDTA) pH 8 and stored at 4°C. 2.5 µl of this cDNA was used in a 25 µl PCR reaction with 12.5 pmol of SBE A and Ro primers for 30 cycles of 94°C 45 sec, 50°C 25 sec, 72°C 1 min 30 sec. A second round of PCR (25 cycles) was performed using 1 µl of this reaction as template in a 50 µl reaction under the same conditions. Amplified products were separated by agarose gel electrophoresis and cloned into the pT7Blue vector (Invitrogen).

For the first round of 5' RACE, 5 µg of total leaf RNA was reverse transcribed as described above using 10 pmol of the SBE II gene specific primer CSBE22. This primer was removed from the reaction by diluting to 500 µl with TE and centrifuging twice through a centricon 100 microconcentrator. The concentrated cDNA was then dA-tailed with 9U of terminal deoxynucleotide transferase and 50 µM dATP in a 20 µl reaction in buffer supplied by the manufacturer (BRL). The reaction was incubated for 10 min at 37°C and 5 min at 65°C and then diluted to 200 µl with TE pH 8. PCR was performed in a 50 µl volume using 5 µl of tailed cDNA, 2.5 pmol of RoRidT17 and 25 pmol of Ro and CSBE24 primers for 30 cycles of 94°C 45 sec, 55°C 25 sec, 72°C 3 min. Amplified products were separated on a 1% TAE agarose gel, cut out, 200 µl of TE was added and melted at 99°C for 10 min. Five µl of this was re-amplified in a 50 µl volume using CSBE25 and Ri as primers and 25 cycles of 94°C 45 sec, 55°C 25 sec, 72°C 1 min 30 sec. Amplified fragments were separated on a 1% TAE agarose gel, purified on DEAE paper and cloned into pT7Blue.

The second round of 5' RACE was performed using CSBE28 and 29 primers in the first and second round PCR reactions respectively using a new A-tailed cDNA library primed

with CSBE27.

A third round of 5' RACE was performed on the same CSBE27 primed cDNA.

Repeat 3' RACE and PCR Cloning

The 3' RACE library (RoRidT17 primed leaf RNA) was used as a template. The first PCR reaction was diluted 1:20 and 1 μ l was used in a 50 μ l PCR reaction with SBE A and Ri primers and the products were cloned into pT7Blue. The cloned PCR products were screened for the presence or absence of the CSBE23 oligo by colony PCR.

A full length cDNA of cassava SBE II was isolated by PCR from leaf or root cDNA (RoRidT17 primed) using primers CSBE214 and CSBE218 from 2.5 μ l of cDNA in a 25 μ l reaction and 30 cycles of 94°C 45 sec, 55°C 25 sec, 72°C 2 min.

Complementation of *E. coli* mutant KVR32

SBE II containing plasmids were transformed into the branching enzyme deficient mutant *E. coli* KVR32 (Keil *et al.*, 1987 Mol. Gen. Genet. 207, 294-301) and cells grown on solid PYG media (0.85 % KH_2PO_4 , 1.1 % K_2HPO_4 , 0.6 % yeast extract) containing 1.0 % glucose. To test for complementation, a loop of cells was scraped off and resuspended in 150 μ L water to which was added 15 μ L of Lugol's solution (2 g KI and 1 g I_2 per 300 ml water).

RNA isolation

RNA was isolated from cassava plants by the method of Logemann (1987 Anal. Biochem. 163, 21-26). Leaf RNA was isolated from 0.5 gm of in vitro grown plant tissue. The total yield was 300 μ g. Three month old roots (88 gm) were used for isolation of root RNA).

SBE II specific oligonucleotides

SBE A	ATGGACAAGGATATGTATGA	(Seq ID No. 1)
CSBE21	GGTTTCATGACTTCTGAGCA	(Seq ID No. 2)

CSBE22	TGCTCAGAAAGTCATGAAACC	(Seq ID No. 3)
CSBE23	TCCAGTCTCAATATACGTCG	(Seq ID No. 4)
CSBE24	AGGAGTAGATGGTCTGTCGA	(Seq ID No. 5)
CSBE25	TCATACATATCCTTGTCAT	(Seq ID No. 6)
CSBE26	GGGTGACTTCAATGATGTAC	(Seq ID No. 7)
CSBE27	GGTGACATCATTGAAGTCA	(Seq ID No. 8)
CSBE28	AATTACTGGCTCCGTACTAC	(Seq ID No. 9)
CSBE29	CATTCCAACGTGCGACTCAT	(Seq ID No. 10)
CSBE210	TACCGGTAATCTAGGTGTTG	(Seq ID No. 11)
CSBE211	GGACCTTGGTTTAGATCAA	(Seq ID No. 12)
CSBE212	ATGAGTCGCACGTTGGAATG	(Seq ID No. 13)
CSBE213	CAACACCTAGATTACCGTA	(Seq ID No. 14)
CSBE214	TTAGTTGCGTCAGTTCTCAC	(Seq ID No. 15)
CSBE215	AATATCTATCTCAGCCGGAG	(Seq ID No. 16)
CSBE216	ATCTTAGATAGTCTGCATCA	(Seq ID No. 17)
CSBE217	TGGTTGTTCCCTGGAATTAC	(Seq ID No. 18)
CSBE218	TGCAAGGACCGTGACATCAA	(Seq ID No. 19)

RESULTS

Cloning of a SBE II gene from cassava leaf

The strategy for cloning a full length cDNA of starch branching enzyme II of cassava is shown in Figure 1. A comparison of several SBE II (class A) SBE DNA sequences identified a 23 bp region which appears to be completely conserved among most genes (data not shown) and is positioned about one kilobase upstream from the 3' end of the gene. An oligonucleotide primer (designated SBE A) was made to this sequence and used to isolate a partial cDNA clone by 3' RACE PCR from first strand leaf cDNA as illustrated in Figure 1. An approximately 1100 bp band was amplified, cloned into pT7Blue vector and sequenced. This clone was designated pSJ94 and contained a 1120 bp insert starting with the SBE A oligo and ending with a polyA tail. There was a predicted open reading frame of 235 amino acids which was highly homologous (79% identical) to a potato SBE II also isolated by the inventors (data not shown) suggesting that this clone represented a class A (SBE II) gene.

To obtain the sequence of a full length clone nested primers were made complementary to the 5' end of this sequence and used in 5' RACE PCR to isolate clones from the 5' region of the gene. A total of three rounds of 5' RACE was needed to determine the sequence of the complete gene (i.e. one that has a predicted long ORF preceded by stop codons). It should be noted that during this cloning process several clones (# 23, 9, 16) were obtained that had small deletions and in one case (clone 23) there was also a small (120 bp) intron present. These occurrences are not uncommon and probably arise through errors in the PCR process and/or reverse transcription of incompletely processed RNA (heterogeneous nuclear RNA).

The overlapping cDNA fragments could be assembled into a contiguous 3 kb sequence (designated csbe2con.seq) which contained one long predicted ORF as shown in Figure 2. Several clones in the last round of 5' RACE were obtained which included sequence of the untranslated leader (UTL). All of these clones had an ORF (42 amino acids) 46 bp upstream and out of frame with that of the long ORF.

There is more than one SBE II gene in cassava

In order to determine if the assembled sequence represented that of a single gene, attempts were made to recover by PCR a full length SBE II gene using primers CSBE214 and CSBE23 at the 5' and 3' ends of the csbe2con sequence respectively. All attempts were unsuccessful using either leaf or root cDNA as template. The PCR was therefore repeated with either the 5' or 3' most primer and complementary primers along the length of the SBE II gene to determine the size of the largest fragment that could be amplified. With the CSBE214 primer, fragments could be amplified using primers 210, 28, 27 and 22 in order of increasing distance, the latter primer pair amplifying a 2.2 kb band. With the 3' primer CSBE23, only primer pairs with 21 and 26 gave amplification products, the latter being about 1200 bp. These results suggest that the original 3' RACE clone (pSJ94) is derived from a different SBE II gene than the rest of the 5' RACE clones even though the two largest PCR fragments (214+22 and 26+23) overlap by 750 bp and share several primer sites. It is likely that the sequence of the two genes starts to diverge around the CSBE22 primer site such that the 3' end of the corresponding gene does not contain the 23 primer and is not therefore able to amplify a cDNA when used with the 214 primer.

To confirm this, the sequence of the longest 5' PCR fragment (214+22) from two clones (#20 designated pSJ99, & #35) was determined and compared to the consensus sequence csbe2con as shown in Figure 3. The first 2000 bases are nearly identical (the single base changes might well be PCR errors), however the consensus sequence is significantly different after this. This region corresponds to the original 3' RACE fragment pSJ94 (SBE A + Ri adaptor) and provided evidence that there may be more than one SBE II gene in cassava.

The 3' end corresponding to pSJ99 was therefore cloned as follows: 3' RACE PCR was performed on leaf cDNA using the SBE A oligo as the gene specific primer so that all SBE II genes would be amplified. The cloned DNA fragments were then screened for the presence or absence of the CSBE23 primer by PCR. Two out of 15 clones were positive with the SBE A + Ri primer pair but negative with SBE A + CSBE23 primers. The sequence of these two clones (designated pSJ101, as shown in Figure 9) demonstrated that they were indeed from an SBE II gene and that they were different from pSJ94. However the overlapping region of pSJ101 (the 3' clone) and pSJ99 (the 5' clone) was identical suggesting that they were derived from the same gene.

To confirm this a primer (CSBE218) was made to a region in the 3' UTR (untranslated region) of pSJ101 and used in combination with CSBE214 primer to recover by PCR a full length cDNA from both leaf and root cDNA. These clones were sequenced and designated pSJ106 & pSJ107 respectively. The sequence and predicted ORF of pSJ107 is shown in Figure 4. The long ORF in plasmid pSJ106 was found to be interrupted by a stop codon (presumably introduced in the PCR process) approximately 1 kb from the 3' end of the gene, therefore another cDNA clone (designated pSJ116) was amplified in a separate reaction, cloned and sequenced. This clone had an intact ORF (data not shown). There were only a few differences in these two sequences (in the transit peptide aa 27-41: YRRTSSCLSFNFKEA to DRRTSSCLSFIFKAA and L831 in pSJ107 to V in pSJ116 respectively).

An additional 740bp of sequence of the gene corresponding to the pSJ94 clone was isolated by 5' RACE using the primers CSBE216 and 217, and was designated pSJ125.

This sequence was combined with that of pSJ94 to form a consensus sequence "125 + 94", as shown in Figure 10. The sequence of this second gene is about 90% identical at the DNA and protein level to pSJ116, as shown in Figure 5 and 6, and is clearly a second form of SBE II in cassava. The 3' untranslated regions of the two genes are not related (data not shown).

It was also determined that the full length cassava SBE II genes (from both leaf and tuber) actually encode for active starch branching enzymes since the cloned genes were able to complement the glycogen branching enzyme deficient *E. coli* mutant KVR32.

Main Findings

1) A full length cDNA clone of a starch branching enzyme II (SBE II) gene has been cloned from leaves and starch storing roots of cassava. This cDNA encodes a 836 amino acid protein (Mr 95 Kd) and is 86 % identical to pea SBE I over the central conserved domain, although the level of sequence identity over the entire coding region is lower than 86%.

2) There is more than one SBE II gene in cassava as a second partial SBE II cDNA was isolated which differs slightly in the protein coding region from the first gene and has no homology in the 3' untranslated region.

3) The isolated full length cDNA from both leaves and roots encodes an active SBE as it complements an *E. coli* mutant deficient in glycogen branching enzyme as assayed by iodine staining.

We have shown that there are SBE II (Class A) gene sequences present in the cassava genome by isolating cDNA fragments using 3' and 5' RACE. From these cDNA fragments a consensus sequence of over 3 kb could be compiled which contained one long open reading frame (Figure 2) which is highly homologous to other SBE II (class A) genes (data not shown). It is likely that the consensus sequence does not represent that of a single gene since attempts to PCR a full length gene using primers at the 5' and 3' ends of this sequence were not successful. In fact screening of a number of leaf derived 3'

RACE cDNAs showed that a second SBE II gene (clone designated pSJ101) was also expressed which is highly homologous within the coding region to the originally isolated cDNA (pSJ94) but has a different 3' UTR. A full length SBE II gene was isolated from leaves and roots by PCR using a new primer to the 3' end of this sequence and the original sequence at the 5' end of the consensus sequence. If the frequency of clones isolated by 3' RACE PCR reflects the abundance of the mRNA levels then this full length gene may be expressed at lower levels in the leaf than the pSJ94 clone (2 out of 15 were the former class, 13/15 the latter). It should be noted that each class is expressed in both leaves and roots as judged by PCR (data not shown). Sequence analysis of the predicted ORF of the leaf and root genes showed only a few differences (4 amino acid changes and one deletion) which could have arisen through PCR errors or, alternatively, there may be more than one nearly identical gene expressed in these tissues.

A comparison of all known SBE II protein sequences shows that the cassava SBE II gene is most closely related to the pea gene (Figure 8). The two proteins are 86.3% identical over a 686 amino acid range which extends from the triple proline "elbow" (Burton *et al.*, 1995 Plant J. 7, 3-15) to the conserved VVYA sequence immediately preceding the C-terminal extensions (data not shown). All SBE II proteins are conserved over this range in that they are at least 80% similar to each other. Remarkably however, the sequence conservation between the pea, potato and cassava SBE II proteins also extends to the N-terminal transit peptide, especially the first 12 amino acids of the precursor protein and the region surrounding the mature terminus of the pea protein (AKFSRDS). Because the proteins are so similar around this region it can be predicted that the mature terminus of the cassava SBE II protein is likely to be GKSSHES. The precursor has a predicted molecular mass of 96 kD and the mature protein a predicted molecule mass of 91.3 kD. The cassava SBE II has a short acidic tail at the C-terminal although this is not as long or as acidic as that found in the pea or potato proteins. The significance of this acidic tail, if any, remains to be determined. One notable difference between the amino acid sequence of cassava SBE II and all other SBE II proteins is the presence of the sequence NSKH at around position 697 instead of the conserved sequence DAD/EY. Although this conserved region forms part of a predicted α -helix (number 8) of the catalytic $(\beta/\alpha)_8$ barrel domain (Burton *et al* 1995 cited previously), this difference does not abolish the SBE

activity of the cassava protein as this gene can still complement the glycogen branching deletion mutant of *E. coli*. It may however affect the specificity of the protein. An interesting point is that the other cassava SBE II clone pSJ94 has the conserved sequence DADY.

One other point of interest concerning the sequence of the SBE II gene is the presence of an upstream ATG in the 5' UTR. This ATG could initiate a small peptide of 42 amino acids which would terminate downstream of the predicted initiating methionine codon of the SBE II precursor. If this does occur then the translation of the SBE II protein from this mRNA is likely to be inefficient as ribosomes normally initiate at the 5' most ATG in the mRNA. However the first ATG is in a poorer Kozak context than the SBE II initiator and it may be too close to the 5' end of the message to initiate efficiently (14 nucleotides) thus allowing initiation to occur at the correct ATG.

In conclusion we have shown that cassava does have SBE II gene sequences, that they are expressed in both leaves and tubers and that more than one gene exists.

Example 2

Cloning of a second full length cassava SBE II gene

Methods

Oligonucleotides

CSBE219	CTTTATCTATTAAAGACTTC	(Seq ID No. 20)
CSBE220	CAAAAAAGTTTGTGACATGG	(Seq ID No. 21)
CSBE221	TCACTTTTCCAATGCTAAT	(Seq ID No. 22)
CSBE222	TTCATGCAATGGAACCGAC	(Seq ID No. 23)
CSBE223	CAGATGTCCTGACTCGAAT	(Seq ID No. 24)
CSBE224	ATTCCGAGTCAAGACATCTG	(Seq ID No. 25)
CSBE225	CGCATTTCTCGCTATTGCTT	(Seq ID No. 26)
CSBE226	CACAGGCCCAAGTGAAGAAT	(Seq ID No. 27)

The 5' end of the gene corresponding to the 3'RACE clone pSJ94 was isolated in three

rounds of 5'RACE. Prior to performing the first round of 5' RACE, 5 µg of total leaf RNA was reverse transcribed in a 20 µl reaction using conditions as described by the manufacturer (Superscript enzyme, BRL) and 10 pmol of the SBE II gene specific primer CSBE23. Primers were then removed and the cDNA tailed with dATP as described above. The first round of 5'RACE used primers CSBE216 and Ro. This PCR reaction was diluted 1:20 and used as a template for a second round of amplification using primers CSBE217 and Ri. The gene specific primers were designed so that they would preferentially hybridise to the SBE II sequence in pSJ94. Amplified products appeared as a smear of approximately 600-1200 bp when subjected to electrophoresis on a 1% TAE agarose gel.

This smear was excised and DNA purified using a Qiaquick column (Qiagen) before ligation to the pT7Blue vector. Several clones were sequenced and clone #7 was designated pSJ125. New primers (CSBE219 and 220) were designed to hybridise to the 5' end of pSJ125 and a second round of 5'RACE was performed using the same CSBE23 primed library. Two fragments of 600 and 800 bp were cloned and sequenced (clones 13,17). Primers CSBE221 and 222 were designed to hybridise to the 5' sequence of the longest clone (#13) and a third round of 5' RACE was performed on a new library (5 µg total leaf RNA reverse transcribed with Superscript using CSBE220 as primer and then dATP tailed with TdT from Boehringer Mannheim). Fragments of approximately 500 bp were amplified, cloned and sequenced. Clone #13, was designated pSJ143. The process is illustrated schematically in Figure 12.

To isolate a full length gene as a contiguous sequence, a new primer (CSBE225) was designed to hybridise to the 5' end of clone pSJ143 and used with one of the primers (CSBE226 or 23) in the 3' end of clone pSJ94, in a PCR reaction using RoRidT17 primed leaf cDNA as template. Use of primer CSBE226 resulted in production of Clone #2 (designated pSJ144), and use of primer CSBE23 resulted in production of Clones #10 and 13 (designated pSJ145 and pSJ146 respectively). Only pSJ146 was sequenced fully.

Results

Isolation of a second full length cassava SBE II gene

A full length clone for a second SBE II gene was isolated by extending the sequence of pSJ94 in three rounds of 5' RACE as illustrated schematically in Figure 12. In each round of 5' RACE, primers were designed that would preferentially hybridise to the new sequence rather than to the gene represented by pSJ116. In the final round of 5' RACE, three clones were obtained that had the initiating methionine codon, and none of these had upstream ATGs. The overlapping cDNA fragments (sequences of the 5' RACE clones pSJ143, 13, pSJ125 and the 3' RACE clone pSJ94) could be assembled into a consensus sequence of approximately 3 kb which was designated *csbe2-2.seq*. This sequence contained one long ORF with a predicted size of 848 aa (M_r 97 kDa). The full length gene was then isolated as a contiguous sequence by PCR amplification from RoRidT17 primed leaf cDNA using primers at the 5' (CSBE225) and 3' (CSBE23 or CSBE226) ends of the RACE clones. One clone, designated pSJ146, was sequenced and the restriction map is shown along with the predicted amino acid sequence in Figure 13.

Sequence homologies between SBE II genes

The two cassava genes (pSJ116 and pSJ146) share 88.8% identity at the DNA level over the entire coding region (data not shown). The homology extends about 50 bases outside of this region but beyond this the untranslated regions show no similarity (data not shown). At the protein level the two genes show 86% identity over the entire ORF (data not shown). The two genes are more closely related to each other than to any other SBE II. Between species, the pea SBE I shows the most homology to the cassava SBE II genes.

Example 3

Construction of plant transformation vectors and transformation of cassava with antisense starch branching enzyme genes.

This example describes in detail how a portion of the SBE II gene isolated from cassava may be introduced into cassava plants to create transgenic plants with altered properties.

An 1100 bp *Hind* III - *Sac* I fragment of cassava SBE II (from plasmid pSJ94) was cloned into the *Hind* III - *Sac* I sites of the plant transformation vector pSJ64 (Figure 11). This placed the SBE II gene in an antisense orientation between the 2X 35S CaMV promoter

- 5 and the nopaline synthase polyadenylation signal. PSJ64 is a derivative of the binary vector pGPTV-HYG (Becker et al., 1992 Plant Molecular Biology 20: 1195-1197) modified by inclusion of an approximately 750 bp fragment of pJIT60 (Guerineau et al 1992 Plant Mol. Biol. 18, 815-818) containing the duplicated cauliflower mosaic virus (CaMV) 35S promoter (Cabb-JI strain, 10 equivalent to nucleotides 7040 to 7376 duplicated upstream of 7040 to 7433, as described by Frank et al., 1980 Cell 21, 285-294) to replace the GUS coding sequence. A similar construct was made with the cassava SBE II sequence from plasmid pSJ101.

- 15 These plasmids are then introduced into *Agrobacterium tumefaciens* LBA4404 by a direct DNA uptake method (An et al. Binary vectors, In: Plant Molecular Biology Manual (ed embryos by selecting on hygromycin as described by Li et al. (1996, Nature Biotechnology 14, 738-740).

- 20 The term "comprise" and grammatical variations thereof such as "comprising" when used in the description or claims does not preclude the presence of additional features, integers, steps or components.

25

30



SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

- (A) NAME: National Starch and Chemical Investment Holding Corporation
- (B) STREET: Suite 27, 501 Silverside Road
- (C) CITY: Wilmington
- (D) STATE: Delaware
- (E) COUNTRY: USA
- (F) POSTAL CODE (ZIP): 19809

(ii) TITLE OF INVENTION: Improvements in or Relating to Starch Content of Plants

(iii) NUMBER OF SEQUENCES: 31 -

(iv) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

ATGGACAAGG ATATGTATGA

20

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

GGTTTCATGA CTTCTGAGCA

20

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

SUBSTITUTE SHEET (RULE 26)

24

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

TGCTCAGAAG TCATGAAACC

20

(2) INFORMATION FOR SEQ ID NO: 4:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

TCCAGTCTCA ATATACGTCG

20

(2) INFORMATION FOR SEQ ID NO: 5:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

AGGAGTAGAT GGTCTGTCCA

20

(2) INFORMATION FOR SEQ ID NO: 6:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

TCATACATAT CCTTGTCAT

20

(2) INFORMATION FOR SEQ ID NO: 7:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

25

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GGGTGACTTC AATGATGTAC

20

(2) INFORMATION FOR SEQ ID NO: 8:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GGGTACATC ATTGAAGTCA

20

(2) INFORMATION FOR SEQ ID NO: 9:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

AATTACTGGC TCGTACTAC

20

(2) INFORMATION FOR SEQ ID NO: 10:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

CATTCCAACG TGCGACTCAT

20

(2) INFORMATION FOR SEQ ID NO: 11:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

TACCGGTAAT CTAGGTGTTG

20

26

(2) INFORMATION FOR SEQ ID NO: 12:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

GGACCTTGGT TTAGATCCAA

20

(2) INFORMATION FOR SEQ ID NO: 13:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

ATGAGTCGCA CGTTGGAATG

20

(2) INFORMATION FOR SEQ ID NO: 14:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

CAACACCTAG ATTACCGTA

20

(2) INFORMATION FOR SEQ ID NO: 15:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

TTAGTTGCGT CAGTTCTCAC

20

(2) INFORMATION FOR SEQ ID NO: 16:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs

27

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

AATATCTATC TCAGCCGGAG

20

(2) INFORMATION FOR SEQ ID NO: 17:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

ATCTTAGATA GTCTGCATCA

20

(2) INFORMATION FOR SEQ ID NO: 18:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

TGGTGTGTC CTGGAATTAC

20

(2) INFORMATION FOR SEQ ID NO: 19:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

TGCAAGGACC GTGACATCAA

20

(2) INFORMATION FOR SEQ ID NO: 20:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

28

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

CTTTATCTAT TAAAGACTTC

20

(2) INFORMATION FOR SEQ ID NO: 21:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

CAAAAAAGTT TGTGACATGG

20

(2) INFORMATION FOR SEQ ID NO: 22:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

TCACTTTTTC CAATGCTAAT

20

(2) INFORMATION FOR SEQ ID NO: 23:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

TTCATGCAA TGAACCGAC

20

(2) INFORMATION FOR SEQ ID NO: 24:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

GAGATGTCCT GACTCGGAAT

20

SUBSTITUTE SHEET (RULE 28)

(2) INFORMATION FOR SEQ ID NO: 25:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

ATTCGAGTC AGGACATCTG

20

(2) INFORMATION FOR SEQ ID NO: 26:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

CGCATTCTC GCTATTGCTT

20

(2) INFORMATION FOR SEQ ID NO: 27:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

CACAGGCCCA AGTGAAGAAT

20

(2) INFORMATION FOR SEQ ID NO: 28:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2588 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
 (B) LOCATION: 21..2531

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

CTCTCTAACT TCTCAGCGAA ATG GGA CAC TAC ACC ATA TCA GGA ATA CGT
 Met Gly His Tyr Thr Ile Ser Gly-Ile Arg
 1 5 10

50

30

TTT CCT TGT GCT CCA CTC TGC AAA TCT CAA TCT ACC GGC TTC CAT GGC Phe Pro Cys Ala Pro Leu Cys Lys Ser Gln Ser Thr Gly Phe His Gly 15 20 25	98
TAT CGG AGG ACC TCC TCT TGC CTT TCC TTC AAC TTC AAG GAG GCG TTT Tyr Arg Arg Thr Ser Ser Cys Leu Ser Phe Asn Phe Lys Glu Ala Phe 30 35 40	146
TCT AGG AGG GTC TTC TCT GGA AAG TCA TCT CAT GAA TCT GAC TCC TCA Ser Arg Arg Val Phe Ser Gly Lys Ser Ser His Glu Ser Asp Ser Ser 45 50 55	194
AAT GTA ATG GTC ACT GCT TCT AAA AGA GTC CTT CCT GAT GGT CGG ATT Asn Val Met Val Thr Ala Ser Lys Arg Val Leu Pro Asp Gly Arg Ile 60 65 70	242
GAA TGC TAT TCT TCT TCA ACA GAT CAA TTG GAA GCC CCT GGC ACA GTT Glu Cys Tyr Ser Ser Ser Thr Asp Gln Leu Glu Ala Pro Gly Thr Val 75 80 85 90	290
TCA GAA GAA TCC CAG GTG CTT ACT GAT GTT GAG AGT CTC ATT ATG GAT Ser Glu Glu Ser Gln Val Leu Thr Asp Val Glu Ser Leu Ile Met Asp 95 100 105	338
GAT AAG ATT GTT GAA GAT GAA GTA AAT AAA GAA TCT GTT CCA ATG CGG Asp Lys Ile Val Glu Asp Glu Val Asn Lys Glu Ser Val Pro Met Arg 110 115 120	386
GAG ACA GTT AGC ATC AGA AAA ATT GGA TCT AAA CCA AGG TCC ATT CCT Glu Thr Val Ser Ile Arg Lys Ile Gly Ser Lys Pro Arg Ser Ile Pro 125 130 135	434
CCA CCC GGC AGA GGG CAA AGA ATA TAT GAC ATA GAT CCA AGC TTG ACA Pro Pro Gly Arg Gly Gln Arg Ile Tyr Asp Ile Asp Pro Ser Leu Thr 140 145 150	482
GGC TTT CGT CAA CAC CTA GAT TAC CGG TAT TCA CAG TAC AAA AGA CTC Gly Phe Arg Gln His Leu Asp Tyr Arg Tyr Ser Gln Tyr Lys Arg Leu 155 160 165 170	530
CGA GAA GAA ATT GAC AAG TAT GAA GGT AGT CTG GAT GCA TTT TCT CGT Arg Glu Glu Ile Asp Lys Tyr Glu Gly Ser Leu Asp Ala Phe Ser Arg 175 180 185	578
GGC TAT GAA AAG TTT GGT TTC TCA CGC AGT GAA ACA GGA ATA ACT TAT Gly Tyr Glu Lys Phe Gly Phe Ser Arg Ser Glu Thr Gly Ile Thr Tyr 190 195 200	626
AGA GAG TGG GCA CCA GGA GCT ACG TGG GCT GCA TTG ATT GGA GAT TTC Arg Glu Trp Ala Pro Gly Ala Thr Trp Ala Ala Leu Ile Gly Asp Phe 205 210 215	674
AAT AAC TGG AAT CCT AAT GCA GAT GTC ATG ACT CAG AAT GAG TGT GGT Asn Asn Trp Asn Pro Asn Ala Asp Val Met Thr Gln Asn Glu Cys Gly 220 225 230	722

31

GTC TGG GAG ATC TTT TTG CCG AAT AAT GCA GAT GGT TCA CCA CCA ATT Val Trp Glu Ile Phe Leu Pro Asn Asn Ala Asp Gly Ser Pro Pro Ile 235 240 245 250	770
CCC CAT GGT TCT CGA GTA AAG ATA CGC ATG GAT ACT CCA TCT GGC AAC Pro His Gly Ser Arg Val Lys Ile Arg Met Asp Thr Pro Ser Gly Asn 255 260 265	818
AAA GAT TCT ATT CCT GCT TGG ATC AAG TTC TCA GTT CAA GCA CCA GGT Lys Asp Ser Ile Pro Ala Trp Ile Lys Phe Ser Val Gln Ala Pro Gly 270 275 280	866
GAA CTC CCA TAT AAT GGC ATA TAC TAT GAT CCT CCC GAG GAG GAG AAG Glu Leu Pro Tyr Asn Gly Ile Tyr Asp Pro Pro Glu Glu Glu Lys 285 290 295	914
TAT GTG TTC AAA AAT CCT CAG CCA AAG AGA CCA AAA TCA CTT CGG ATT Tyr Val Phe Lys Asn Pro Gln Pro Lys Arg Pro Lys Ser Leu Arg Ile 300 305 310	962
TAT GAG TCG CAC GTT GGA ATG AGT AGT ACG GAG CCA GTA ATT AAC ACA Tyr Glu Ser His Val Gly Met Ser Ser Thr Glu Pro Val Ile Asn Thr 315 320 325 330	1010
TAT GCC AAC TTT AGA GAT GAT GTG CTT CCT CGC ATC AAA AAG CTT GGC Tyr Ala Asn Phe Arg Asp Asp Val Leu Pro Arg Ile Lys Lys Leu Gly 335 340 345	1058
TAC AAT GCT GTT CAG CTC ATG GCT ATT CAA GAG CAT TCA TAT TAT GCT Tyr Asn Ala Val Gln Leu Met Ala Ile Gln Glu His Ser Tyr Tyr Ala 350 355 360	1106
AGT TTT GGG TAT CAC GTC ACA AAC TTT TAT GCA GCT AGC AGC CGA TTT Ser Phe Gly Tyr His Val Thr Asn Phe Tyr Ala Ala Ser Ser Arg Phe 365 370 375	1154
GGA ACT CCT GAT GAT TTA AAG TCT CTA ATA GAT AAA GCT CAC GAG TTA Gly Thr Pro Asp Asp Leu Lys Ser Leu Ile Asp Lys Ala His Glu Leu 380 385 390	1202
GGT CTT CTT GTT CTC ATG GAT ATT GTT CAT AGC CAT GCA TCA ACT AAT Gly Leu Leu Val Leu Met Asp Ile Val His Ser His Ala Ser Thr Asn 395 400 405 410	1250
ACG TTG GAT GGG CTG AAT ATG TTT GAT GGT ACG GAT GGT CAC TAC TTT Thr Leu Asp Gly Leu Asn Met Phe Asp Gly Thr Asp Gly His Tyr Phe 415 420 425	1298
CAC TCT GGA CCA CGG GGT CAT CAT TGG ATG TGG GAC TCT CGC CTT TTC His Ser Gly Pro Arg Gly His His Trp Met Trp Asp Ser Arg Leu Phe 430 435 440	1346
AAC TAT GGG AGC TGG GAG GTT CTA AGG TTT CTT CTT TCA AAT GCA AGG Asn Tyr Gly Ser Trp Glu Val Leu Arg Phe Leu Leu Ser Asn Ala Arg 445 450 455	1394

TGG Trp 460	TGG Trp	TTG Leu	GAT Asp	GAG Glu	TAC Tyr	AAG Lys 465	TTT Phe	GAT Asp	GGG Gly	TTC Phe 470	AGA Arg	TTT Phe	GAT Asp	GGG Gly	GTG Val	1442
ACT Thr 475	TCA Ser	ATG Met	ATG Met	TAC Tyr	ACC Thr 480	CAT His	CAT His	GGA Gly	TTG Leu 485	CAG Gln	GTA Val	GAT Asp	TTT Phe	ACC Thr	GGC Gly 490	1490
AAC Asn	TAC Tyr	AAT Asn	GAA Glu	TAC Tyr 495	TTT Phe	GGA Gly	TAT Tyr	GCA Ala	ACT Thr 500	GAT Asp	GTA Val	GAT Asp	GCT Ala	GTG Val 505	GTT Val	1538
TAT Tyr	TTG Leu	ATG Met	CTG Leu	TTG Leu	AAT Asn 510	GAT Asp	ATG Met	ATT Ile 515	CAT His	GGT Gly	CTC Leu	TTC Phe	CCA Pro	GAG Glu	GCT Ala	1586
GTC Val	ACC Thr	ATT Ile	GGT Gly	GAA Glu	GAT Asp	GTT Val 525	AGT Ser	GGA Gly	ATG Met	CCA Pro	ACA Thr	GTT Val	TGC Cys	ATT Ile	CGG Pro	1634
GTT Val	GAA Glu	GAT Asp	GGT Gly	GGT Gly	GTT Val 540	GGC Gly	TTT Phe	GAT Asp	TAT Tyr	CGT Arg	CTC Leu	CAC His	ATG Met	GCT Ala	GTT Val	1682
GCT Ala 555	GAT Asp	AAA Lys	TGG Trp	GTT Val	GAG Glu 560	ATT Ile	ATT Ile	CAG Gln	AAG Lys	AGA Arg 565	GAT Asp	GAA Glu	GAT Asp	TGG Trp	AAA Lys 570	1730
ATG Met	GGT Gly	GAC Asp	ATT Ile	GTA Val	CAT His	ATG Met	CTG Leu	ACC Thr	AAC Asn	AGG Arg 580	CGG Arg	TGG Trp	TTG Leu	GAA Glu	AAG Lys 585	1778
TGT Cys	GTT Val	TCT Ser	TAT Tyr	GCT Ala	GAA Glu	AGT Ser	CAT His	GAC Asp	CAG Gln	GCC Ala	CTT Leu	GTT Val	GGT Gly	GAC Asp	AAA Lys 600	1826
ACT Thr	ATT Ile	GCA Ala	TTT Phe	TGG Trp	CTG Leu	ATG Met	GAC Asp	AAG Lys	GAT Asp	ATG Met	TAT Tyr	GAC Asp	TTC Phe	ATG Met	GCT Ala	1874
CTT Leu	GAC Asp	AGA Arg	CCA Pro	TCT Ser	ACT Thr	CCT Pro	CTC Leu	ATA Ile	GAT Asp	CGT Arg	GGA Gly	GTA Val	GCA Ala	TTG Leu	CAC His	1922
AAA Lys 635	ATG Met	ATC Ile	AGG Arg	CTT Leu	ATT Ile	ACC Thr	ATG Met	GGA Gly	TTA Leu	GGC Gly	GGA Gly	GAA Glu	GGA Gly	TAT Tyr	TTG Leu 650	1970
AAT Asn	TTT Phe	ATG Met	GGA Gly	AAT Asn	GAA Glu	TTT Phe	GGA Gly	CAC His	CCC Pro	GAG Glu	TGG Trp	ATT Ile	GAT Asp	TTT Phe	CCA Pro	2018
AGA Arg	GGT Gly	GAT Asp	CTA Leu	CAT His	CTT Leu	CCC Pro	AGT Ser	GGT Gly	AAA Lys	TTT Phe	GTT Val	CCT Pro	GGG Gly	AAC Asn	AAT Asn	2066

33

TAC AGT TAT GAT AAA TGC CGG CGT AGG TTT GAT CTA GGC AAT TCA AAG Tyr Ser Tyr Asp Lys Cys Arg Arg Arg Phe Asp Leu Gly Asn Ser Lys 685 690 695	2114
CAT CTG AGA TAT CAT GGA ATG CAA GAG TTT GAT CAA GCA ATT CAG CAT His Leu Arg Tyr His Gly Met Gln Glu Phe Asp Gln Ala Ile Gln His 700 705 710	2162
CTT GAA GAA GCC TAT GGT TTC ATG ACT TCT GAG CAC CAA TAC ATA TCA Leu Glu Glu Ala Tyr Gly Phe Met Thr Ser Glu His Gln Tyr Ile Ser 715 720 725 730	2210
CGG AAG GAT GAA AGG GAT CGG ATC ATT GTC TTC GAG AGG GGA AAC CTC Arg Lys Asp Glu Arg Asp Arg Ile Ile Val Phe Glu Arg Gly Asn Leu 735 740 745	2258
GTT TTT GTA TTC AAT TTT CAT TGG ACT AGC AGC TAT TCG GAT TAC CGA Val Phe Val Phe Asn Phe His Trp Thr Ser Ser Tyr Ser Asp Tyr Arg 750 755 760	2306
GTT GGC TGC TTA AAG CCA GGA AAG TAC AAG ATA GTC TTG GAT TCA GAT Val Gly Cys Leu Lys Pro Gly Lys Tyr Lys Ile Val Leu Asp Ser Asp 765 770 775	2354
GAT CCT TTG TTT GGA GGC TTT GGC AGG CTT AGT CAT GAT GCA GAG CAC Asp Pro Leu Phe Gly Gly Phe Gly Arg Leu Ser His Asp Ala Glu His 780 785 790	2402
TTT AGC TTT GAA GGG TGG TAC GAT AAC CGG CCT CGA TCC TTC ATG GTG Phe Ser Phe Glu Gly Trp Tyr Asp Asn Arg Pro Arg Ser Phe Met Val 795 800 805 810	2450
TAC ACA CCA TGT AGA ACA GCA GTG GTC TAT GCT TTA GTG GAG GAT GAA Tyr Thr Pro Cys Arg Thr Ala Val Val Tyr Ala Leu Val Glu Asp Glu 815 820 825	2498
GTG GAG AAT GAA TTG GAA CCT GTC GCC GGT TAA GATATATCTT AACACAGGT Val Glu Asn Glu Leu Glu Pro Val Ala Gly * 830 835	2551
TCTGAAGCAG GAATGCCATT ATTGATCTTC CTATGTT	2588

(2) INFORMATION FOR SEQ ID NO: 29:

- (i) SEQUENCE CHARACTERISTICS;
 (A) LENGTH: 837 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

Met Gly His Tyr Thr Ile Ser Gly Ile Arg Phe Pro Cys Ala Pro Leu
 1 5 10 15

Cys Lys Ser Gln Ser Thr Gly Phe His Gly Tyr Arg Arg Thr Ser Ser
 20 25 30
 Cys Leu Ser Phe Asn Phe Lys Glu Ala Phe Ser Arg Arg Val Phe Ser
 35 40 45
 Gly Lys Ser Ser His Glu Ser Asp Ser Ser Asn Val Met Val Thr Ala
 50 55 60
 Ser Lys Arg Val Leu Pro Asp Gly Arg Ile Glu Cys Tyr Ser Ser Ser
 65 70 75 80
 Thr Asp Gln Leu Glu Ala Pro Gly Thr Val Ser Glu Glu Ser Gln Val
 85 90 95
 Leu Thr Asp Val Glu Ser Leu Ile Met Asp Asp Lys Ile Val Glu Asp
 100 105 110
 Glu Val Asn Lys Glu Ser Val Pro Met Arg Glu Thr Val Ser Ile Arg
 115 120 125
 Lys Ile Gly Ser Lys Pro Arg Ser Ile Pro Pro Pro Gly Arg Gly Gln
 130 135 140
 Arg Ile Tyr Asp Ile Asp Pro Ser Leu Thr Gly Phe Arg Gln His Leu
 145 150 155 160
 Asp Tyr Arg Tyr Ser Gln Tyr Lys Arg Leu Arg Glu Glu Ile Asp Lys
 165 170 175
 Tyr Glu Gly Ser Leu Asp Ala Phe Ser Arg Gly Tyr Glu Lys Phe Gly
 180 185 190
 Phe Ser Arg Ser Glu Thr Gly Ile Thr Tyr Arg Glu Trp Ala Pro Gly
 195 200 205
 Ala Thr Trp Ala Ala Leu Ile Gly Asp Phe Asn Asn Trp Asn Pro Asn
 210 215 220
 Ala Asp Val Met Thr Gln Asn Glu Cys Gly Val Trp Glu Ile Phe Leu
 225 230 235 240
 Pro Asn Asn Ala Asp Gly Ser Pro Pro Ile Pro His Gly Ser Arg Val
 245 250 255
 Lys Ile Arg Met Asp Thr Pro Ser Gly Asn Lys Asp Ser Ile Pro Ala
 260 265 270
 Trp Ile Lys Phe Ser Val Gln Ala Pro Gly Glu Leu Pro Tyr Asn Gly
 275 280 285
 Ile Tyr Tyr Asp Pro Pro Glu Glu Glu Lys Tyr Val Phe Lys Asn Pro
 290 295 300
 Gln Pro Lys Arg Pro Lys Ser Leu Arg Ile Tyr Glu Ser His Val Gly
 305 310 315 320

35

Met Ser Ser Thr Glu Pro Val Ile Asn Thr Tyr Ala Asn Phe Arg Asp
 325 330 335
 Asp Val Leu Pro Arg Ile Lys Lys Leu Gly Tyr Asn Ala Val Gln Leu
 340 345 350
 Met Ala Ile Gln Glu His Ser Tyr Tyr Ala Ser Phe Gly Tyr His Val
 355 360 365
 Thr Asn Phe Tyr Ala Ala Ser Ser Arg Phe Gly Thr Pro Asp Asp Leu
 370 375 380
 Lys Ser Leu Ile Asp Lys Ala His Glu Leu Gly Leu Leu Val Leu Met
 385 390 395 400
 Asp Ile Val His Ser His Ala Ser Thr Asn Thr Leu Asp Gly Leu Asn
 405 410 415
 Met Phe Asp Gly Thr Asp Gly His Tyr Phe His Ser Gly Pro Arg Gly
 420 425 430
 His His Trp Met Trp Asp Ser Arg Leu Phe Asn Tyr Gly Ser Trp Glu
 435 440 445
 Val Leu Arg Phe Leu Leu Ser Asn Ala Arg Trp Trp Leu Asp Glu Tyr
 450 455 460
 Lys Phe Asp Gly Phe Arg Phe Asp Gly Val Thr Ser Met Met Tyr Thr
 465 470 475 480
 His His Gly Leu Gln Val Asp Phe Thr Gly Asn Tyr Asn Glu Tyr Phe
 485 490 495
 Gly Tyr Ala Thr Asp Val Asp Ala Val Val Tyr Leu Met Leu Leu Asn
 500 505 510
 Asp Met Ile His Gly Leu Phe Pro Glu Ala Val Thr Ile Gly Glu Asp
 515 520 525
 Val Ser Gly Met Pro Thr Val Cys Ile Pro Val Glu Asp Gly Gly Val
 530 535 540
 Gly Phe Asp Tyr Arg Leu His Met Ala Val Ala Asp Lys Trp Val Glu
 545 550 555 560
 Ile Ile Gln Lys Arg Asp Glu Asp Trp Lys Met Gly Asp Ile Val His
 565 570 575
 Met Leu Thr Asn Arg Arg Trp Leu Glu Lys Cys Val Ser Tyr Ala Glu
 580 585 590
 Ser His Asp Gln Ala Leu Val Gly Asp Lys Thr Ile Ala Phe Trp Leu
 595 600 605
 Met Asp Lys Asp Met Tyr Asp Phe Met Ala Leu Asp Arg Pro Ser Thr
 610 615 620

SUBSTITUTE SHEET (RULE 28)

36

Pro Leu Ile Asp Arg Gly Val Ala Leu His Lys Met Ile Arg Leu Ile
 625 630 635 640
 Thr Met Gly Leu Gly Gly Glu Gly Tyr Leu Asn Phe Met Gly Asn Glu
 645 650 655
 Phe Gly His Pro Glu Trp Ile Asp Phe Pro Arg Gly Asp Leu His Leu
 660 665 670
 Pro Ser Gly Lys Phe Val Pro Gly Asn Asn Tyr Ser Tyr Asp Lys Cys
 675 680 685
 Arg Arg Arg Phe Asp Leu Gly Asn Ser Lys His Leu Arg Tyr His Gly
 690 695 700
 Met Gln Glu Phe Asp Gln Ala Ile Gln His Leu Glu Glu Ala Tyr Gly
 705 710 715 720
 Phe Met Thr Ser Glu His Gln Tyr Ile Ser Arg Lys Asp Glu Arg Asp
 725 730 735
 Arg Ile Ile Val Phe Glu Arg Gly Asn Leu Val Phe Val Phe Asn Phe
 740 745 750
 His Trp Thr Ser Ser Tyr Ser Asp Tyr Arg Val Gly Cys Leu Lys Pro
 755 760 765
 Gly Lys Tyr Lys Ile Val Leu Asp Ser Asp Asp Pro Leu Phe Gly Gly
 770 775 780
 Phe Gly Arg Leu Ser His Asp Ala Glu His Phe Ser Phe Glu Gly Trp
 785 790 795 800
 Tyr Asp Asn Arg Pro Arg Ser Phe Met Val Tyr Thr Pro Cys Arg Thr
 805 810 815
 Ala Val Val Tyr Ala Leu Val Glu Asp Glu Val Glu Asn Glu Leu Glu
 820 825 830
 Pro Val Ala Gly *
 835

(2) INFORMATION FOR SEQ ID NO: 30:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2805 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(1x) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 131..2677

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

AGTGAATTCG AGCTCGGTAC CCGGGGATCC GATTGSCATT TCTCGCTATT GCTTTCCGTT	60
TATTTCCATA TATAAAATAT CAAATCTAAT CACTTGGGCC ATTTCTATCT CTCTCCAAAC	120
TCTCACCGAA ATG GTA TAC TAC ACT GTA TCA GGC ATA CGT TTT CCT TGT Met Val Tyr Tyr Thr Val Ser Gly Ile Arg Phe Pro Cys 840 845 850	169
GCA CCT TCA CTC TAC AAA TCT CAG CTC ACC AGC TTC CAT GGC GGT CGA Ala Pro Ser Leu Tyr Lys Ser Gln Leu Thr Ser Phe His Gly Gly Arg 855 860 865	217
AGG ACC TCT TCT GGC CTT TCC TTC CTC TTG AAG AAG GAG CTG TTT CCT Arg Thr Ser Ser Gly Leu Ser Phe Leu Leu Lys Lys Glu Leu Phe Pro 870 875 880	265
CGG AAG ATC TTT GCT GGA AAG TCC TCT TAT GAA TCT GAC TCC TCA AAT Arg Lys Ile Phe Ala Gly Lys Ser Ser Tyr Glu Ser Asp Ser Ser Asn 885 890 895	313
TTA ACT GTC TCT GCA TCT GAG AAG GTC CTT GTT CCT GAT GAT CAG ATT Leu Thr Val Ser Ala Ser Glu Lys Val Leu Val Pro Asp Asp Gln Ile 900 905 910	361
GAT GGC TCT TCT TCT TCA ACA TAT CAA TTA GAA ACC ACT GGC ACA GTT Asp Gly Ser Ser Ser Ser Thr Tyr Gln Leu Glu Thr Thr Gly Thr Val 915 920 925 930	409
TTG GAG GAA TCC CAG GTT CTT GGT GAT GCA GAG AGT CTT GTG ATG GAA Leu Glu Glu Ser Gln Val Leu Gly Asp Ala Glu Ser Leu Val Met Glu 935 940 945	457
GAT GAT AAG AAT GTT GAG GAG GAT GAA GTA AAA AAA GAG TCG GTT CCA Asp Asp Lys Asn Val Glu Glu Asp Glu Val Lys Lys Glu Ser Val Pro 950 955 960	505
TTG CAT GAG ACA ATT AGC ATT GGA AAA AGT GAA TCT AAA CCA AGG TCC Leu His Glu Thr Ile Ser Ile Gly Lys Ser Glu Ser Lys Pro Arg Ser 965 970 975	553
ATT CCT CCA CCT GGC AGT GGG CAG AGA ATA TAT GAC ATA GAT CCA AGC Ile Pro Pro Pro Gly Ser Gly Gln Arg Ile Tyr Asp Ile Asp Pro Ser 980 985 990	601
TTG GCA GGT TTC CGT CAG CAT CTT GAC TAC CGA TAT TCA CAG TAC AAA Leu Ala Gly Phe Arg Gln His Leu Asp Tyr Arg Tyr Ser Gln Tyr Lys 995 1000 1005 1010	649
AGG CTG CGT GAG GAA ATT GAC AAG TAT GAA GGT GGT TTG GAT GCA TTC Arg Leu Arg Glu Glu Ile Asp Lys Tyr Glu Gly Gly Leu Asp Ala Phe 1015 1020 1025	697
TCT CGT GGA TTT GAA AAG TTT GGT TTC TTA CGC AGT GAA ACA GGA ATA Ser Arg Gly Phe Glu Lys Phe Gly Phe Leu Arg Ser Glu Thr Gly Ile 1030 1035 1040	745

ACT TAT AGG GAA TGG GCA CCT GGA GCT ACG TGG GCT GCA CTT ATT GGA Thr Tyr Arg Glu Trp Ala Pro Gly Ala Thr Trp Ala Ala Leu Ile Gly 1045 1050 1055	793
GAT TTC AAC AAT TGG AAT CCT AAT GCA GAT GTC ATG ACT CGG AAT GAG Asp Phe Asn Asn Trp Asn Pro Asn Ala Asp Val Met Thr Arg Asn Glu 1060 1065 1070	841
TTT GGT GTC TGG GAG ATT TTT TTG CCA AAT AAC GCA GAT GGT TCA CCA Phe Gly Val Trp Glu Ile Phe Leu Pro Asn Asn Ala Asp Gly Ser Pro 1075 1080 1085 1090	889
CCA ATT CCT CAT GGT TCT CGA GTA AAG ATA CGC ATG GAT ACT CCA TCT Pro Ile Pro His Gly Ser Arg Val Lys Ile Arg Met Asp Thr Pro Ser 1095 1100 1105	937
GGC ATC AAA GAT TCA ATT CCT GCT TGG ATC AAG TTC TCA GTT CAG GCA Gly Ile Lys Asp Ser Ile Pro Ala Trp Ile Lys Phe Ser Val Gln Ala 1110 1115 1120	985
CCT GGT GAA ATC CCA TAC AAT GCC ATA TAC TAT GAT CCA CCA AAG GAG Pro Gly Glu Ile Pro Tyr Asn Ala Ile Tyr Tyr Asp Pro Pro Lys Glu 1125 1130 1135	1033
GAG AAG TAT GTG TTC AAA CAT CCT CAG CCA AAG-AGA CCA AAA TCA CTT Glu Lys Tyr Val Phe Lys His Pro Gln Pro Lys Arg Pro Lys Ser Leu 1140 1145 1150	1081
AGG ATT TAT GAA TCT CAT GTT GGG ATG AGT AGT ATG GAG CCA ATA ATT Arg Ile Tyr Glu Ser His Val Gly Met Ser Ser Met Glu Pro Ile Ile 1155 1160 1165 1170	1129
AAC ACA TAT GCC AAC TTT AGA GAT GAT ATG CTT CCT CGC ATC AAA AAG Asn Thr Tyr Ala Asn Phe Arg Asp Asp Met Leu Pro Arg Ile Lys Lys 1175 1180 1185	1177
CTT GGC TAC AAT GCT GTT CAG ATC ATG GCT ATT CAA GAG CAT TCC TAT Leu Gly Tyr Asn Ala Val Gln Ile Met Ala Ile Gln Glu His Ser Tyr 1190 1195 1200	1225
TAT GCT AGT TTT GGG TAC CAT GTC ACA AAC TTT TTT GCA CCT AGC AGC Tyr Ala Ser Phe Gly Tyr His Val Thr Asn Phe Phe Ala Pro Ser Ser 1205 1210 1215	1273
CGA TTT GGA ACT CCT GAT GAT TTG AAG TCT TTA ATA GAT AAA GCT CAT Arg Phe Gly Thr Pro Asp Asp Leu Lys Ser Leu Ile Asp Lys Ala His 1220 1225 1230	1321
GAG TTA GGG CTG CTT GTT CTC ATG GAT ATT GTT CAT AGC CAT GCG TCA Glu Leu Gly Leu Leu Val Leu Met Asp Ile Val His Ser His Ala Ser 1235 1240 1245 1250	1369
AAT AAT ACG TTG GAT GGG CTG AAC ATG TTT GAT GGT ACG GAT AGT CAC Asn Asn Thr Leu Asp Gly Leu Asn Met Phe Asp Gly Thr Asp Ser His 1255 1260 1265	1417

TAC TTC CAC TCC GGA TCA CGG GGT CAT CAT TGG TTG TGG GAC TCT CGC Tyr Phe His Ser Gly Ser Arg Gly His His Trp Leu Trp Asp Ser Arg 1270 1275 1280	1465
CTT TTC AAC TAT GGA AGC TGG GAG GTG CTA AGA TTT CTT CTT TCA AAT Leu Phe Asn Tyr Gly Ser Trp Glu Val Leu Arg Phe Leu Leu Ser Asn 1285 1290 1295	1513
GCA AGA TGG TGG TTG GAA GAG TAC AGG TTT GAT GGT TTT AGA TTT GAT Ala Arg Trp Trp Leu Glu Glu Tyr Arg Phe Asp Gly Phe Arg Phe Asp 1300 1305 1310	1561
GGG GTG ACT TCC ATG ATG TAC ACT CCC CAT GGG TTG CAG GTA GCT TTT Gly Val Thr Ser Met Met Tyr Thr Pro His Gly Leu Gln Val Ala Phe 1315 1320 1325 1330	1609
ACT GGC AAC TAC AAT GAG TAC TTT GGA TAT GCA ACT GAT GTA GAT GCT Thr Gly Asn Tyr Asn Glu Tyr Phe Gly Tyr Ala Thr Asp Val Asp Ala 1335 1340 1345	1657
GTG ATT TAT TTG ATG CTT GTG AAT GAT ATG ATT CAG GGT CTT TTC CCT Val Ile Tyr Leu Met Leu Val Asn Asp Met Ile His Gly Leu Phe Pro 1350 1355 1360	1705
GAG GCT GTT ACC ATT GGT GAA GAT GTT AGC GGA AAG CCA ACA TTT TGC Glu Ala Val Thr Ile Gly Glu Asp Val Ser Gly Lys Pro Thr Phe Cys 1365 1370 1375	1753
ATT CCA GTG GAA GAT GGT GGT GTT GGA TTT GAT TAC CGT CTC CAC ATG Ile Pro Val Glu Asp Gly Gly Val Gly Phe Asp Tyr Arg Leu His Met 1380 1385 1390	1801
GCC ATT GCC GAT AAA TGG ATT GAG ATT CTT AAG AAG AGA GAT GAG GAC Ala Ile Ala Asp Lys Trp Ile Glu Ile Leu Lys Lys Arg Asp Glu Asp 1395 1400 1405 1410	1849
TGG AAA ATG GGT GAC ATT GTG CAT ACA CTC ACC AAC AGA AGG TGG TTG Trp Lys Met Gly Asp Ile Val His Thr Leu Thr Asn Arg Arg Trp Leu 1415 1420 1425	1897
GAA AAA TGT GTT GCT TAT GCT GAA AGT CAT GAC CAA GCT CTT GTT GGT Glu Lys Cys Val Ala Tyr Ala Glu Ser His Asp Gln Ala Leu Val Gly 1430 1435 1440	1945
GAC AAA ACT ATT GCA TTT TGG CTG ATG GAC AAG GAC ATG TAC GAC TTC Asp Lys Thr Ile Ala Phe Trp Leu Met Asp Lys Asp Met Tyr Asp Phe 1445 1450 1455	1993
ATG GCT CGT GAC AGA CCA TCT ACT CCT CTT ATA GAT CGT GGA ATA GCA Met Ala Arg Asp Arg Pro Ser Thr Pro Leu Ile Asp Arg Gly Ile Ala 1460 1465 1470	2041
TTG CAC AAA ATG ATC AGG CTT ATT ACC ATG GGC TTA GGC GGA GAA GGA Leu His Lys Met Ile Arg Leu Ile Thr Met Gly Leu Gly Gly Glu Gly 1475 1480 1485 1490	2089

TAT TTG AAT TTT ATG GGA AAT GAA TTT GGA CAT CCT GAG TGG ATT GAT Tyr Leu Asn Phe Met Gly Asn Glu Phe Gly His Pro Glu Trp Ile Asp 1495 1500 1505	2137
TTT CCA AGA GGG GAT CGA CAT CTG CCC AAT GGT AAA GTA ATT CCA GGG Phe Pro Arg Gly Asp Arg His Leu Pro Asn Gly Lys Val Ile Pro Gly 1510 1515 1520	2185
AAC AAC CAC AGT TAT GAT AAA TGC CGT CGT AGA TTT GAT CTA GGT GAT Asn Asn His Ser Tyr Asp Lys Cys Arg Arg Arg Phe Asp Leu Gly Asp 1525 1530 1535	2233
GCA GAC TAT CTA AGA TAT CAT GGA ATG CAA GAG TTT GAT CAG GCA ATG Ala Asp Tyr Leu Arg Tyr His Gly Met Gln Glu Phe Asp Gln Ala Met 1540 1545 1550	2281
CAA CAT CTT GAA GAA GCC TAT GGT TTC ATG ACT TCT GAG CAC CAG TAT Gln His Leu Glu Glu Ala Tyr Gly Phe Met Thr Ser Glu His Gln Tyr 1555 1560 1565 1570	2329
ATA TCA CGG AAG GAT GAA GGA GAT CGG ATC ATT GTC TTT GAG AGG GGA Ile Ser Arg Lys Asp Glu Gly Asp Arg Ile Ile Val Phe Glu Arg Gly 1575 1580 1585	2377
AAC CTT GTT TTT GTA TTC AAC TTT CAT TGG ACT AAC AGC TAT TCA GAT Asn Leu Val Phe Val Phe Asn Phe His Trp Thr Asn Ser Tyr Ser Asp 1590 1595 1600	2425
TAC CGA GTT GGC TGC TTC AAG TCA GGA AAG TAC AAG ATT GTT TTG GAC Tyr Arg Val Gly Cys Phe Lys Ser Gly Lys Tyr Lys Ile Val Leu Asp 1605 1610 1615	2473
TCG GAT GAT GGC TTG TTT GGA GGC TTC AAC AGG CTT AGT CAT GAT GCC Ser Asp Asp Gly Leu Phe Gly Gly Phe Asn Arg Leu Ser His Asp Ala 1620 1625 1630	2521
GAG CAC TTC ACC TTT GAC GGG TGG TAT GAT AAC CGG CCT CGG TCC TTC Glu His Phe Thr Phe Asp Gly Trp Tyr Asp Asn Arg Pro Arg Ser Phe 1635 1640 1645 1650	2569
ATG GTA TAT GCA CCA TCT AGG ACA GCA GTG GTC TAT GCT TTA GTA GAA Met Val Tyr Ala Pro Ser Arg Thr Ala Val Val Tyr Ala Leu Val Glu 1655 1660 1665	2617
GAT GAA GAG AAT GAA GCA GAG AAT GAA GTA GAA AGT GAA GTG AAA CCA Asp Glu Glu Asn Glu Ala Glu Asn Glu Val Glu Ser Glu Val Lys Pro 1670 1675 1680	2665
GCC TCC GGC TGA GATAGATATT TAGTAAGAGG ATCCCCTAAA GCAGGAATGG Ala Ser Gly * 1685	2717
TTAACCTGTG CATCTGCATT GAACGACGTA TATTGAGACT GGAAATCCAT ATGACTAGTA GATCCTCTAG AGTCGACCTG CAGGCATG	2777 2805

(2) INFORMATION FOR SEQ ID NO: 31:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 849 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

```

Met Val Tyr Tyr Thr Val Ser Gly Ile Arg Phe Pro Cys Ala Pro Ser
 1           5           10
Leu Tyr Lys Ser Gln Leu Thr Ser Phe His Gly Gly Arg Arg Thr Ser
 20          25          30
Ser Gly Leu Ser Phe Leu Leu Lys Lys Glu Leu Phe Pro Arg Lys Ile
 35          40          45
Phe Ala Gly Lys Ser Ser Tyr Glu Ser Asp Ser Ser Asn Leu Thr Val
 50          55          60
Ser Ala Ser Glu Lys Val Leu Val Pro Asp Asp Gln Ile Asp Gly Ser
 65          70          75
Ser Ser Ser Thr Tyr Gln Leu Glu Thr Thr Gly Thr Val Leu Glu Glu
 85          90          95
Ser Gln Val Leu Gly Asp Ala Glu Ser Leu Val Met Glu Asp Asp Lys
100         105         110
Asn Val Glu Glu Asp Glu Val Lys Lys Glu Ser Val Pro Leu His Glu
115         120         125
Thr Ile Ser Ile Gly Lys Ser Glu Ser Lys Pro Arg Ser Ile Pro Pro
130         135         140
Pro Gly Ser Gly Gln Arg Ile Tyr Asp Ile Asp Pro Ser Leu Ala Gly
145         150         155         160
Phe Arg Gln His Leu Asp Tyr Arg Tyr Ser Gln Tyr Lys Arg Leu Arg
165         170         175
Glu Glu Ile Asp Lys Tyr Glu Gly Gly Leu Asp Ala Phe Ser Arg Gly
180         185         190
Phe Glu Lys Phe Gly Phe Leu Arg Ser Glu Thr Gly Ile Thr Tyr Arg
195         200         205
Glu Trp Ala Pro Gly Ala Thr Trp Ala Ala Leu Ile Gly Asp Phe Asn
210         215         220
Asn Trp Asn Pro Asn Ala Asp Val Met Thr Arg Asn Glu Phe Gly Val
225         230         235         240

```

Trp Glu Ile Phe Leu Pro Asn Asn Ala Asp Gly Ser Pro Pro Ile Pro
 245 250 255
 His Gly Ser Arg Val Lys Ile Arg Met Asp Thr Pro Ser Gly Ile Lys
 260 265 270
 Asp Ser Ile Pro Ala Trp Ile Lys Phe Ser Val Gln Ala Pro Gly Glu
 275 280 285
 Ile Pro Tyr Asn Ala Ile Tyr Tyr Asp Pro Pro Lys Glu Glu Lys Tyr
 290 295 300
 Val Phe Lys His Pro Gln Pro Lys Arg Pro Lys Ser Leu Arg Ile Tyr
 305 310 315 320
 Glu Ser His Val Gly Met Ser Ser Met Glu Pro Ile Ile Asn Thr Tyr
 325 330 335
 Ala Asn Phe Arg Asp Asp Met Leu Pro Arg Ile Lys Lys Leu Gly Tyr
 340 345 350
 Asn Ala Val Gln Ile Met Ala Ile Gln Glu His Ser Tyr Tyr Ala Ser
 355 360 365
 Phe Gly Tyr His Val Thr Asn Phe Phe Ala Pro Ser Ser Arg Phe Gly
 370 375 380
 Thr Pro Asp Asp Leu Lys Ser Leu Ile Asp Lys Ala His Glu Leu Gly
 385 390 395 400
 Leu Leu Val Leu Met Asp Ile Val His Ser His Ala Ser Asn Asn Thr
 405 410 415
 Leu Asp Gly Leu Asn Met Phe Asp Gly Thr Asp Ser His Tyr Phe His
 420 425 430
 Ser Gly Ser Arg Gly His His Trp Leu Trp Asp Ser Arg Leu Phe Asn
 435 440 445
 Tyr Gly Ser Trp Glu Val Leu Arg Phe Leu Leu Ser Asn Ala Arg Trp
 450 455 460
 Trp Leu Glu Glu Tyr Arg Phe Asp Gly Phe Arg Phe Asp Gly Val Thr
 465 470 475 480
 Ser Met Met Tyr Thr Pro His Gly Leu Gln Val Ala Phe Thr Gly Asn
 485 490 495
 Tyr Asn Glu Tyr Phe Gly Tyr Ala Thr Asp Val Asp Ala Val Ile Tyr
 500 505 510
 Leu Met Leu Val Asn Asp Met Ile His Gly Leu Phe Pro Glu Ala Val
 515 520 525
 Thr Ile Gly Glu Asp Val Ser Gly Lys Pro Thr Phe Cys Ile Pro Val
 530 535 540

Glu Asp Gly Gly Val Gly Phe Asp Tyr Arg Leu His Met Ala Ile Ala
 545 550 555 560
 Asp Lys Trp Ile Glu Ile Leu Lys Lys Arg Asp Glu Asp Trp Lys Met
 565 570 575
 Gly Asp Ile Val His Thr Leu Thr Asn Arg Arg Trp Leu Glu Lys Cys
 580 585 590
 Val Ala Tyr Ala Glu Ser His Asp Gln Ala Leu Val Gly Asp Lys Thr
 595 600 605
 Ile Ala Phe Trp Leu Met Asp Lys Asp Met Tyr Asp Phe Met Ala Arg
 610 615 620
 Asp Arg Pro Ser Thr Pro Leu Ile Asp Arg Gly Ile Ala Leu His Lys
 625 630 635 640
 Met Ile Arg Leu Ile Thr Met Gly Leu Gly Gly Glu Gly Tyr Leu Asn
 645 650 655
 Phe Met Gly Asn Glu Phe Gly His Pro Glu Trp Ile Asp Phe Pro Arg
 660 665 670
 Gly Asp Arg His Leu Pro Asn Gly Lys Val Ile Pro Gly Asn Asn His
 675 680 685
 Ser Tyr Asp Lys Cys Arg Arg Arg Phe Asp Leu Gly Asp Ala Asp Tyr
 690 695 700
 Leu Arg Tyr His Gly Met Gln Glu Phe Asp Gln Ala Met Gln His Leu
 705 710 715 720
 Glu Glu Ala Tyr Gly Phe Met Thr Ser Glu His Gln Tyr Ile Ser Arg
 725 730 735
 Lys Asp Glu Gly Asp Arg Ile Ile Val Phe Glu Arg Gly Asn Leu Val
 740 745 750
 Phe Val Phe Asn Phe His Trp Thr Asn Ser Tyr Ser Asp Tyr Arg Val
 755 760 765
 Gly Cys Phe Lys Ser Gly Lys Tyr Lys Ile Val Leu Asp Ser Asp Asp
 770 775 780
 Gly Leu Phe Gly Gly Phe Asn Arg Leu Ser His Asp Ala Glu His Phe
 785 790 795 800
 Thr Phe Asp Gly Trp Tyr Asp Asn Arg Pro Arg Ser Phe Met Val Tyr
 805 810 815
 Ala Pro Ser Arg Thr Ala Val Val Tyr Ala Leu Val Glu Asp Glu Glu
 820 825 830
 Asn Glu Ala Glu Asn Glu Val Glu Ser Glu Val Lys Pro Ala Ser Gly
 835 840 845 *

CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:

1. A nucleic acid sequence encoding a polypeptide, the encoded polypeptide comprising at least a portion of the amino acid sequence shown in Figure 4 or Figure 13 which retains sufficient starch branching enzyme (SBE) activity when expressed in *E. coli* KV 832 to complement the branching enzyme mutation therein; at least 200 bp; and exhibiting at least 88% sequence identity with the corresponding region of the DNA sequence shown in Figures 4, 9, 10 or 13, operably linked in the sense or anti-sense orientation to a promoter operable in plants.
2. A nucleic acid sequence according to claim 1, comprising nucleotides 21-2531 of the nucleic acid sequence shown in Figure 4, or a functionally equivalent nucleotide sequence which hybridises under stringent hybridisation conditions with the nucleic acid sequence shown in Figure 4.
3. A nucleic acid sequence according to claim 1, comprising nucleotides 131-2677 of the nucleic acid sequence shown in Figure 13, or a functionally equivalent sequence which hybridises under stringent hybridisation conditions with the nucleic acid sequence shown in Figure 13.
4. A nucleic acid sequence according to any one of claims 1, 2 or 3 comprising a 5' and/or a 3' untranslated region.
5. A nucleic acid sequence according to any one of claims 1 to 4, encoding a polypeptide having the amino acid sequence NSKH at about residue 697.
6. A nucleic acid sequence comprising at least 200 bp and exhibiting at least 88% sequence identity with the corresponding region of the DNA sequence shown in Figures 4, 9, 10 or 13, operably linked in the sense or anti-sense orientation to a promoter operable in plants.
7. A nucleic acid sequence according to claim 6, comprising at least 300-600bp.



8. A sequence according to claim 6 or 7, comprising a 5' and/or 3' untranslated region.
- 5 9. A sequence according to claim 8, comprising nucleotides 688-1044 of the sequence shown in figure 9, and/or nucleotides 1507-1800 of the sequence shown in Figure 10.
- 10 10. A sequence according to claim 8, comprising the nucleotide sequence shown in Figure 10.
11. A replicable nucleic acid construct comprising a nucleic acid sequence according to any one of the preceding claims.
- 15 12. A polypeptide comprising at least a portion of the amino acid sequence shown in Figure 4 or Figure 13 which retains sufficient starch branching enzyme (SBE) activity when expressed in *E. coli* KV 832 to complement the branching enzyme mutation therein; encoded by a nucleic acid with at least 200 bp which exhibits at least 88% sequence identity with the corresponding region of the DNA sequence shown in Figures 4, 9, 10 or 13, operably linked in the sense or anti-sense orientation to a promoter operable in plants.
- 20 13. A polypeptide according to claim 12, in substantial isolation from other polypeptides.
- 25 14. A polypeptide according to claim 12 or 13, having the amino acid sequence NSKH at about position 697.
- 30 15. A method of modifying starch *in vitro*, the method comprising treating starch to be modified under suitable conditions with an effective amount of a polypeptide according to any one of claims 12, 13 or 14.
- 35 16. A method of altering a plant host cell, the method comprising introducing into the cell a nucleic acid sequence comprising at least 200 bp and exhibiting at least 88% sequence identity with the corresponding region of the DNA sequence shown in figures 4, 9, 10 or 13, operably linked in the sense or anti-sense orientation to a suitable promoter active in the host cell, and causing transcription of



5 the introduced nucleotide sequence, said transcript and/or the translation product thereof being sufficient to interfere with the expression of a homologous gene naturally present in the host cell, which homologous gene encodes a polypeptide having SBE activity.

10 17. A method according to claim 16, wherein the host cell is from a cassava, banana, potato, pea, tomato, maize, wheat, barley, oat, sweet potato or rice plant.

15 18. A method according to claim 16 or 17, comprising the introduction of one or more further nucleic acid sequences, operably linked in the sense or anti-sense orientation to a suitable promoter active in the host cell, and causing transcription of the one or more further nucleic acid sequences, said transcripts and/or translation products thereof being sufficient to interfere with the expression of homologous gene(s) present in the host cell.

20 19. A method according to claim 18, wherein the one or more further nucleic acid sequences interfere with the expression of a gene involved in starch biosynthesis.

20 20. A method according to claim 18 or 19, wherein the further nucleic acid sequence comprises at least part of an SBE I gene.

25 21. A method according to claim 20, wherein the further nucleic acid sequence comprises at least part of the cassava SBE I gene.

30 22. A method according to any one of claims 16-21, wherein the host cell is selected from one of the following: cassava, banana, potato, pea, tomato, maize, wheat, barley, oat, sweet potato or rice.

35 23. A method according to any one of claims 16-22, wherein the altered host cell gives rise to starch having different properties compared to starch from an unaltered cell.

24. A method according to any one of claims 16-23, further comprising the step of growing the altered host cell into a plant or plantlet.



5 25. A method of obtaining starch having altered properties, comprising growing a plant from an altered host cell according to the method of claim 24, and extracting the starch therefrom.

10 26. A plant or plant cell into which has been artificially introduced a nucleic acid sequence comprising at least 200bp and exhibiting at least 88% sequence identity with the corresponding region of the DNA sequence shown in Figures 4, 9, 10, or 13, operably linked in the sense or anti-sense orientation to a promoter operable in plants, or the progeny thereof.

15 27. A plant according to claim 24, altered by the method of any one of claims 16-22.

20 28. Starch obtainable from an altered plant according to claim 26 or 27, having altered properties compared to starch extracted from an equivalent but unaltered plant.

29. Starch obtained from an altered plant according to claim 26 or 27, having altered properties compared to starch extracted from an equivalent but unaltered plant.

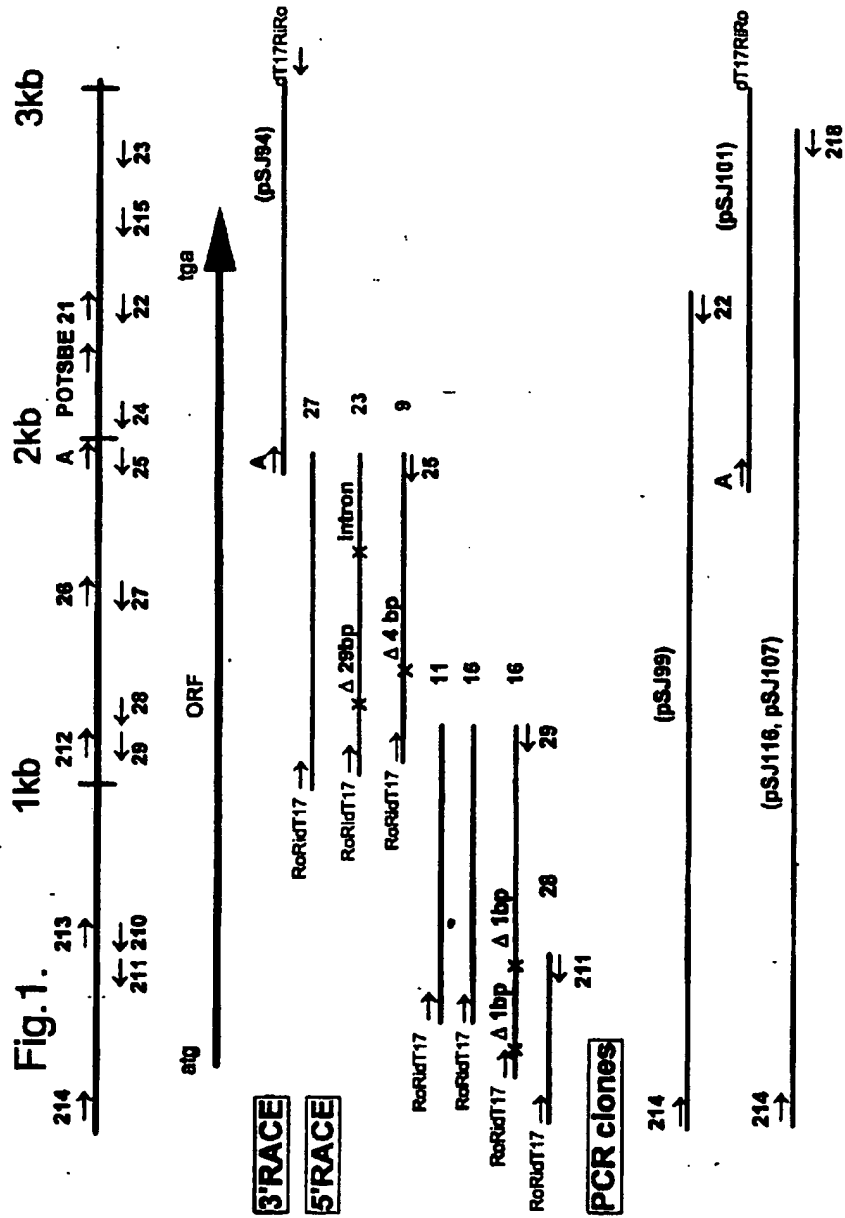
25 30. Starch according to claim 28 or 29 obtained from an altered plant selected from the group consisting of: cassava, banana, potato, pea, tomato, maize, wheat, barley, oat, sweet potato and rice plants.

30 31. Starch according to any one of claims 28, 29 or 30, having increased amylose content compared to starch extracted from an equivalent but unaltered plant.

Dated this 21st day of December 2000

35 NATIONAL STARCH AND CHEMICAL INVESTMENT
HOLDING CORPORATION
By their Patent Attorneys
COLLISON & CO





SUBSTITUTE SHEET (RULE 26)

SUBSTITUTE SHEET (RULE 26)

4/18

Figure 3 displays a grid of 100 small plots, arranged in 10 rows and 10 columns. Each plot shows a time series of data points (represented by small circles) and a fitted curve (represented by a solid line). The plots are labeled with numbers 1 through 100. The data points are scattered across the plots, and the fitted curves vary in shape and position, illustrating the results of the method described in the text. The plots show various patterns of data distribution and curve fitting, illustrating the results of the method described in the text.

Fig. 3.

SUBSTITUTE SHEET (RULE 26)

6/18

Fig.4 (Cont).

[illegible]

7/18

Fig.5.

125-94. seq	^80	^70	^80	^90	^100	^110	^120
116. seq	^1140	^1150	^1180	^1170	^1180	^1190	^1200
125-94. seq	^130	^140	^150	^160	^170	^180	^180
116. seq	^1210	^1220	^1230	^1240	^1250	^1260	^1270
125-94. seq	^200	^210	^220	^230	^240	^250	^260
116. seq	^1280	^1290	^1300	^1310	^1320	^1330	^1340
125-94. seq	^270	^280	^290	^300	^310	^320	^330
116. seq	^1350	^1360	^1370	^1380	^1390	^1400	^1410
125-94. seq	^340	^350	^360	^370	^380	^390	^400
116. seq	^1420	^1430	^1440	^1450	^1460	^1470	^1480
125-94. seq	^410	^420	^430	^440	^450	^460	^470
116. seq	^1490	^1500	^1510	^1520	^1530	^1540	^1550
125-94. seq	^480	^490	^500	^510	^520	^530	^540
116. seq	^1560	^1570	^1580	^1590	^1600	^1610	^1620
125-94. seq	^550	^560	^570	^580	^590	^600	^610
116. seq	^1630	^1640	^1650	^1660	^1670	^1680	^1690
125-94. seq	^620	^630	^640	^650	^660	^670	^680
116. seq	^1700	^1710	^1720	^1730	^1740	^1750	^1760
125-94. seq	^690	^700	^710	^720	^730	^740	^750
116. seq	^1770	^1780	^1790	^1800	^1810	^1820	^1830
125-94. seq	^760	^770	^780	^790	^800	^810	^820
116. seq	^1840	^1850	^1860	^1870	^1880	^1890	^1900
125-94. seq	^830	^840	^850	^860	^870	^880	^890
116. seq	^1910	^1920	^1930	^1940	^1950	^1960	^1970
125-94. seq	^900	^910	^920	^930	^940	^950	^960
116. seq	^1980	^1990	^2000	^2010	^2020	^2030	^2040

Fig.5 (Cont).

	970	980	990	1000	1010	1020	1030
125-94. seq	CCAAGAGGGGATCGACATCTGCCCAATGGTAAAGTAATTCAGGGAACAACCACAGTTATGATAAATGCC						
116. seq	CCAAGAGG SATC ACATCT CCCA TGCTAAA T TTCC GGAACAA ACAGTTATGATAAATGCC						
	2050	2060	2070	2080	2090	2100	2110
125-94. seq	GTCTAGATTTGATCTAGGTGATGCAGACTATCTAAGATATCATGGAATGCAAGASTTTGATCAGGCAAT						
116. seq	G CCGTAG TTTGATCTAGG AT CA A ATCT AGATATCATGGAATGCAAGASTTTGATCA GCAAT						
	2120	2130	2140	2150	2160	2170	2180
125-94. seq	GCAACATCTTGAAGAAGCCTATGCTTTCATGACTTCTGAGCAACAGTATATACGGAAGGATGAAGGA						
116. seq	CA CATCTTGAAGAAGCCTATGCTTTCATGACTTCTGAGCAACAGTATATACGGAAGGATGAAG G						
	2190	2200	2210	2220	2230	2240	2250
125-94. seq	GATCGGATCATTGTCTTTGAGAGGGGAAACCTTGTTTTTGTATTCAACTTTTCATTGGACTAACAGCTATT						
116. seq	GATCGGATCATTGTCTT GAGAGGGGAAACCT GTTTTTGTATTCAA TTTTCATTGGACTA CAGCTATT						
	2260	2270	2280	2290	2300	2310	2320
125-94. seq	CAGATTACCGAGTTGGCTGCTTAAAGTCAGGAAAGTACAAGATTGTTTTGGAAGTCCGATCATGCTTCTT						
116. seq	C GATTACCGAGTTGGCTGCTTAAAGTCAGGAAAGTACAAGATTGTTTTGGAAGTCCGATCATGCTTCTT						
	2330	2340	2350	2360	2370	2380	2390
125-94. seq	TGGAGGCTTCAACAGGCTTAGTCATGATGCCGAGCAGCTTACCTTTGACGGGTGGTATGATAACCGGCTT						
116. seq	TGGAGGCTT CAGGCTTAGTCATGATGC GAGCAGCTTCA CTTTGA GGGTGGTA GATAACCGGCTT						
	2400	2410	2420	2430	2440	2450	2460
125-94. seq	CGGTCTTCATGCTATATGCAACATCTAGGACAGCAGTGGTCCATGCTTTAGTAGAAGATGAAG						
116. seq	CG TCTTCATGCT TA CACCAT TAG ACAGCAGTGGT ATGCTTTAGT CA GATGAAG						
	2470	2480	2490	2500	2510	2520	2530
	CGATCTTCATGCTGTACACACCATGTAGAACAGCAGTGGTCTATGCTTTAGTGGAGGATGAAG						

9/18

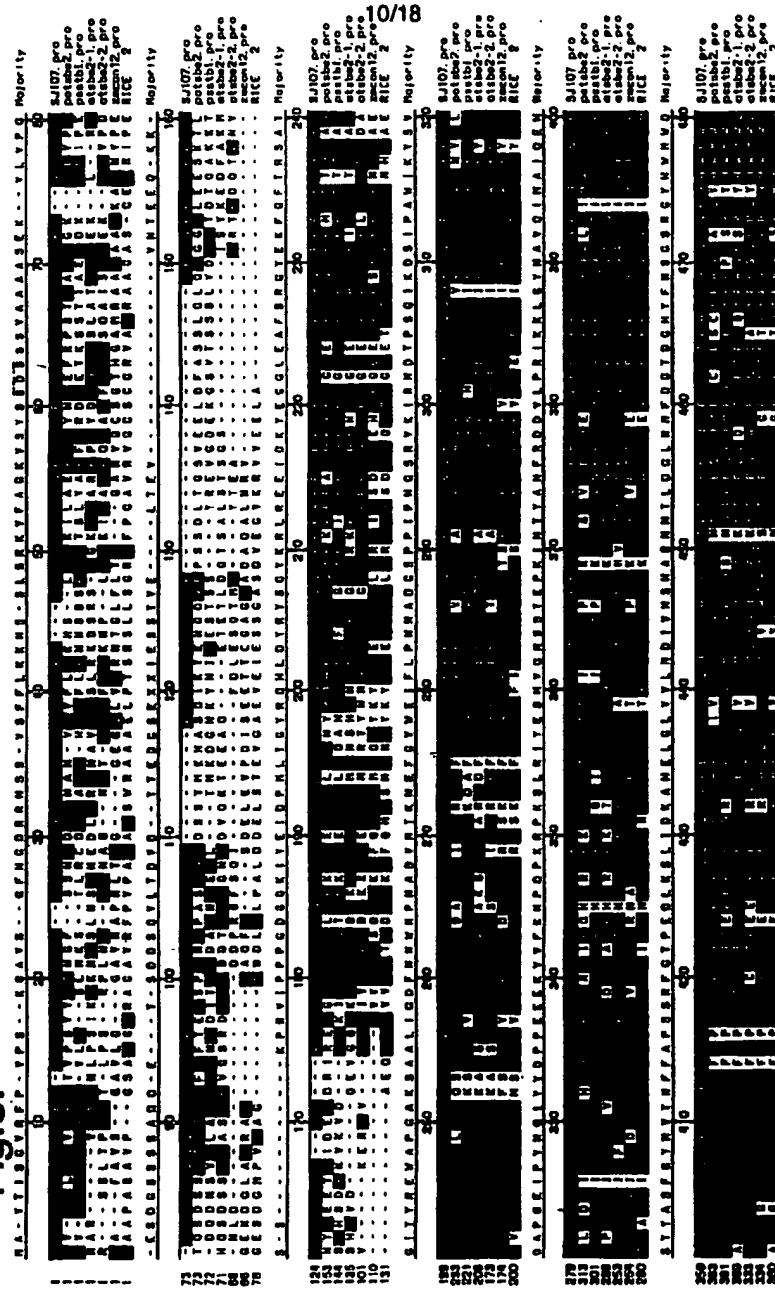
Fig.6.

125-94. pro SFGYHYTNFFAPSSRFCTPDLKSLIDKAEHLGLLVLMIDIVHSHASNTLDGLNMFQDGTDSHYFHSQSRG
 SFGYHYTNF: A: SSRFGTPOOLKSLIDKAEHLGLLVLMIDIVHSHAS. NTLDGLNMFQDGT: HYFHSQ: RG
 116. pro SFGYHYTNFYAASSRFCTPDLKSLIDKAEHLGLLVLMIDIVHSHASTNTLDGLNMFQDGTDSHYFHSQSRG
 125-94. pro HMMWDSRLFNYGSEVLRFLLSNARVWLEEFYRFGFRFQDVTSHMYTPHGLQVAFATNYMEYFGYATDV
 HMMWDSRLFNYGSEVLRFLLSNARVWLE: EY: FQFRFQDVTSHMYT. HGLQV. FTYMEYFGYATDV
 116. pro HMMWDSRLFNYGSEVLRFLLSNARVWLETKFQFRFQDVTSHMYTHHGLQVDFATNYMEYFGYATDV
 125-94. pro DAVIYLYNDNHLGLFPEAVTIGEDYSGKPTFCIPYEDGGVGFQDRLHMAIADKWIEILKKRDEDMKMG
 DAY: YLHL: NDNHLGLFPEAVTIGEDYSG. PT: CIPYEDGGVGFQDRLHMA: ADKW: EIL: KKRDEDMKMG
 116. pro DAVVYLYNDNHLGLFPEAVTIGEDYSGKPTVCIPYEDGGVGFQDRLHMAIADKWIEILKKRDEDMKMG
 125-94. pro DIVMTLTNRRLKCYAYAESDQALYGDKTIAFWLHOKOMYDFHARDRPTPLIDRGIALMKHMLRLITH
 DIVH LTNRRWLEKCY: YAESDQALYGDKTIAFWLHOKOMYDFHARDRPTPLIDRG: ALMKHMLRLITH
 116. pro DIVMTLTNRRLKCYAYAESDQALYGDKTIAFWLHOKOMYDFHARDRPTPLIDRGIALMKHMLRLITH
 125-94. pro GLCGEGLNFMGNEFSGHPEWIDFPRGDRHLPNGKYIPGNNSYDKCRRRFDLGDADYLRTHGMOEFDDAM
 GLCGEGLNFMGNEFSGHPEWIDFPRGDRHLP: SK: PGNNSYDKCRRRFDLG: ...LRTHGMOEFDDAM
 116. pro GLCGEGLNFMGNEFSGHPEWIDFPRGDRHLPNGKYIPGNNSYDKCRRRFDLGNKHLRYHNGEFDQAI
 125-94. pro OHLEEAYGHTSEHQYISRKDEGDRIIYFERGMLYVFNFHWTNSDYRYGCFKSKYKIVLDSDDGLF
 OHLEEAYGHTSEHQYISRKDE: DRIIYFERGMLYVFNFHWT: SYDYRYGCF: K: SKYKIVLDSDD: LF
 116. pro OHLEEAYGHTSEHQYISRKDERDRIIYFERGMLYVFNFHWTSSDYRYGCLPKSKYKIVLDSDDPLF
 125-94. pro GGFNRLSHDAEHFTFGVYDNRPRSFHMYAPSRATVYHALVEDEENEAEYVES
 GGF: RLSDAEHF: F: GMYDNRPRSFHMY: P: RTAVY. ALVEDE: : : V: :
 116. pro GGFNRLSHDAEHFTFGVYDNRPRSFHMYTPCRTAVYALVEDEENEAEYVPAG

Fig.7.



Fig.8.



11/18

Fig.8 (Cont).



Fig.9.

SUBSTITUTE SHEET (RULE 26)

SUBSTITUTE SHEET (RULE 26)

SUBSTITUTE SHEET (RULE 26)

Fig.11.

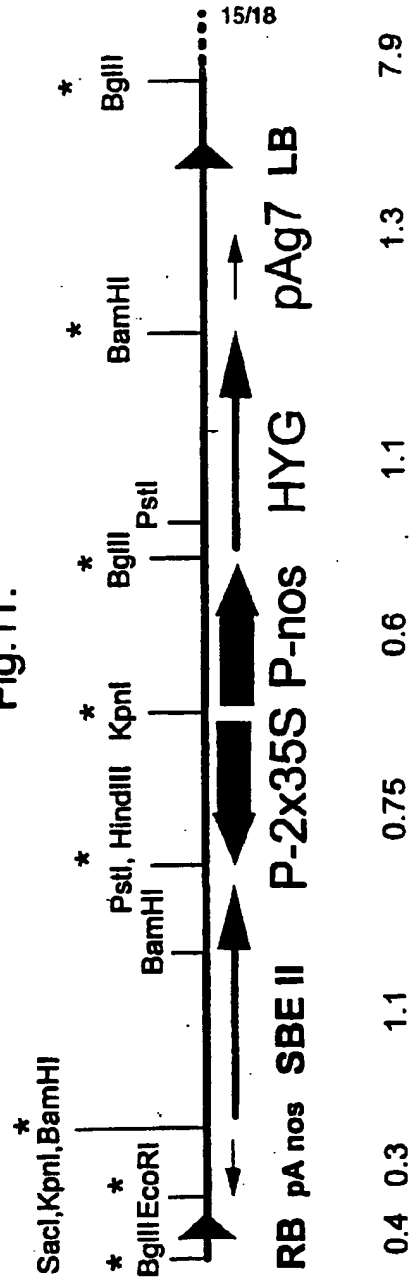


Fig.13.

SUBSTITUTE SHEET (RULE 26)

18/18

Fig.13 (Cont).

[illegible]